

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: February 16, 2005, 19:06:53 ; Search time 170 Seconds
(without alignments)
1758.622 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGRIASRKFKNLGNKGES.....ATVLDKNNISKSTNNPRK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1dDec04:*

- 1: geneeqp19808:*
- 2: geneeqp19908:*
- 3: geneeqp20008:*
- 4: geneeqp20018:*
- 5: geneeqp20028:*
- 6: geneeqp20038:*
- 7: geneeqp20038:*
- 8: geneeqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	4 AAB48343	Aab48343 S. pneumo
2	4026	100.0	2140	6 ABU01020	Abu01020 S. pneumo
3	4026	100.0	2140	6 ABU45746	Abu45746 Protein e
4	4026	100.0	2140	8 ADM92113	Adm92113 S. pneumo
5	4015.5	99.7	2120	3 AAY81710	Aay81710 Streptoco
6	3789.5	94.1	2138	8 ADK48759	Adk48759 Streptoco
7	3048	75.7	637	8 ADR94534	Adr94534 Streptoco
8	750.5	18.6	1529	8 ADR96136	Adr96136 Novel S.
9	615	15.3	117	5 AAW55096	Aaw55096 Streptoco
10	615	15.3	117	5 ABP54590	Abp54590 S. pneumo
11	615	15.3	117	7 ADC45149	Adc45149 S. pneumo
12	227	5.6	1196	6 ABT24813	Abt24813 Protein e
13	222	5.5	861	7 ABQ23508	Abq23508 Plasmodiu
14	220	5.5	1166	8 ADR99186	Adr99186 Streptoco
15	218	5.4	1639	2 AAW54145	Aaw54145 P. falcip
16	218	5.4	1639	5 AAE29345	Aae29345 Plasmodiu
17	217	5.4	1166	8 ADR99185	Adr99185 Streptoco
18	217	5.4	1166	8 ADR99178	Adr99178 Streptoco
19	217	5.4	1166	8 ADR99184	Adr99184 Streptoco
20	216	5.4	1166	8 ADR99183	Adr99183 Streptoco
21	216	5.4	1233	5 ABP30203	Abp30203 Streptoco
22	216	5.4	1233	5 ABP29675	Abp29675 Streptoco
23	216	5.4	1233	8 ADR99176	Adr99176 Streptoco
24	216	5.4	1233	8 ADO10469	Ado10469 Group B S
25	216	5.4	1239	5 ABP25822	Abp25822 Streptoco

ALIGNMENTS

26	215.5	5.4	1254	2	AAR07503	Aar07503 Merozite
27	215.5	5.4	1254	2	AAW24575	Aaw24575 Merozite
28	215	5.3	1166	8	ADR99179	Adr99179 Streptococ
29	212.5	5.3	2485	3	AAB18172	Aab18172 Plasmodiu
30	212	5.3	1141	6	ABU42327	Abu42327 Protein e
31	212	5.3	1188	3	AAB18183	Aab18183 Plasmodiu
32	211.5	5.3	991	8	ADP86452	Adp86452 Clostridiu
33	211.5	5.3	1104	7	ADG73658	Adg73658 C. perfrin
34	209	5.2	1558	3	AAB18324	Aab18324 Plasmodiu
35	208.5	5.2	3029	8	ADP25433	Adp25433 Plasmodiu
36	207.5	5.2	4688	6	ABU48941	Abu48941 Protein e
37	207	5.1	1191	6	ABU24124	Abu24124 Protein e
38	207	5.1	1654	1	AAP50777	Aap50777 Sequence
39	207	5.1	1979	3	AAB18171	Aab18171 Plasmodiu
40	205	5.1	2024	8	ADP25444	Adp25444 Plasmodiu
41	204.5	5.1	1516	3	AAB18195	Aab18195 Plasmodiu
42	204.5	5.1	2500	3	AAB18272	Aab18272 plasmodiu
43	204	5.1	1166	6	AAB18982	Aab18982 Pathogen
44	204	5.1	1245	3	AAB18244	Aab18244 Plasmodiu
45	203	5.0	1166	2	AAY08643	Aay08643 S. aureus

RESULT 1
AAB48343
ID AAB48343 standard; protein; 773 AA.

AC AAB48343;
DT 20-APR-2001 (first entry)
XX S. pneumoniae Spi130 polypeptide.
DB S. pneumoniae Spi130 polypeptide.
XX Immunogenic; Spi128; Spi130; pneumococcal; otitis media; nasopharyngeal;
KW bronchial; lung; blood; infection; immune response; immunotherapy;
KW antibacterial; auditory; vaccine.
XX Streptococcus pneumoniae.
OS Streptococcus pneumoniae.
XX WO200076540-A2.
XX 21-DEC-2000.
XX 09-JUN-2000; 2000MO-US015925.
XX 10-JUN-1999; 99US-0138453P.
XX (MED1-) MED IMMUNE INC.
XX Adamou JE, Choi GH;
XX WPI: 2001-112197/12.
XX N-PSDB; AAC84742.
XX New vaccines comprising Spi128 or Spi130 polypeptides, for treating and
XX preventing pneumococcal infections, particularly infections caused by
XX Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
XX blood infections.
XX Claim 8; Page 51-54; 54pp; English.
XX The invention relates to novel immunogenic polypeptides, Spi128 and Spi130
XX from S. pneumoniae. Vaccines comprising the polypeptides are useful for
XX the treatment and prevention of pneumococcal infections, particularly
XX infections caused by Streptococcus, such as otitis media, nasopharyngeal,
XX bronchial, lung or blood infections. The antigens are used as immunogenic
XX agents to stimulate an immune response. The antisera and antibodies may
XX also be used in diagnosing and treating pneumococcal infections.
XX Recombinant polypeptides serve as a mechanism for stimulating production
XX of antibodies for use in passive immunotherapy, diagnostic reagents, and
XX as reagents in other processes such as affinity chromatography. The

CC present sequence represents the *S. pneumoniae* Sp130 polypeptide
XX
SQ Sequence 773 AA;

Query Match	100.0%	Score 4026;	DB 4;	Length 773;
Best Local Similarly	100.0%	Pred. No. 7.5e-225;		
Matches 773; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	ILGELIABEKFNULGNGBESLKKODTGVHHHONERSIIEKSSPTIDRNIISTIRDEPK	60
Dh	1	KLGBIABESFKNULGNGBESLKKODTGVHHHONERSIIEKSSPTIDRNIISTIRDEPK	60
Qy	61	DLKULIKKKKFEVDDPTSETGKMBEYDYKDDKGNIIAYDDGDTLEFEYIEKULDEISKI	120
Dh	61	DLKULIKKKKFEVDDPTSETGKMBEYDYKDDKGNIIAYDDGDTLEFEYIEKULDEISKI	120
Qy	121	YGVLSPSKDGHEPILIGKISNVSNAKAVYVGNNYKSIIEIKATKYDDEHSKTYTPOLYANIND	180
Dh	121	YGVLSPSKDGHEPILIGKISNVSNAKAVYVGNNYKSIIEIKATKYDDEHSKTYTPOLYANIND	180
Qy	181	IYDGLAFPGDMELPYKXNDOKKAEIKIIMPEKIKETKSEYPIVYSSYGNVIEIIEBGDLSKN	240
Dh	181	IYDGLAFPGDMELPYKXNDOKKAEIKIIMPEKIKETKSEYPIVYSSYGNVIEIIEBGDLSKN	240
Qy	241	KPDNLTKMBSGKIEYDSERQOYLAKONIILRKGYALKVTYVNGKIDMLEGNGVYSKEDI	300
Dh	241	KPDNLTKMBSGKIEYDSERQOYLAKONIILRKGYALKVTYVNGKIDMLEGNGVYSKEDI	300
Qy	301	AKIQANENLRLSETTIYADSRNVEDGSRSTOSVLSALDGFNIIRYQVTFPKANDGGA	360
Dh	301	AKIQANENLRLSETTIYADSRNVEDGSRSTOSVLSALDGFNIIRYQVTFPKANDGGA	360
Qy	361	IDKOGNLTJDSKLYLFGKODKERYGEDEKPNYKAIKEDSGMFLIDTPKPVLSMDKRYFNP	420
Dh	361	IDKOGNLTJDSKLYLFGKODKERYGEDEKPNYKAIKEDSGMFLIDTPKPVLSMDKRYFNP	420
Qy	421	SKSNKIYYRNPBEFYLRGKISDKGFPWBLRVNYESVDNLYIYGDLIHIDTRDYNIKLNVK	480
Dh	421	SKSNKIYYRNPBEFYLRGKISDKGFPWBLRVNYESVDNLYIYGDLIHIDTRDYNIKLNVK	480
Qy	481	DGDIIMDGAKDYKANGPDDKVTJDMOGNVLIQGYSDLNKAVGVHYOPLDYNVKEPVNID	540
Dh	481	DGDIIMDGAKDYKANGPDDKVTJDMOGNVLIQGYSDLNKAVGVHYOPLDYNVKEPVNID	540
Qy	541	PKGNTSIEYADGKSVVFNINDRRNGPDEIIOBHYYINGKETSFPNDIKOIIDKTLNIK	600
Dh	541	PKGNTSIEYADGKSVVFNINDRRNGPDEIIOBHYYINGKETSFPNDIKOIIDKTLNIK	600
Qy	601	IYVNDOPARNTTYKBEFLIANDTGEVSELKPHVTVTIOUNGKMSSTIYSEBDFLLPYKGE	660
Dh	601	IYVNDOPARNTTYKBEFLIANDTGEVSELKPHVTVTIOUNGKMSSTIYSEBDFLLPYKGE	660
Qy	661	LEKGYQPDGWEISGFEKGDAGVYVNLISKOTFLIKPYFKIKBEKKEBKNKPTPVSSKKKON	720
Dh	661	LEKGYQPDGWEISGFEKGDAGVYVNLISKOTFLIKPYFKIKBEKKEBKNKPTPVSSKKKON	720
Qy	721	POVNHQSOLNESHKREDLOREBSHQSDSTFKDYATATYLDNONISSKSTNNPNK 773	
Dh	721	POVNHQSOLNESHKREDLOREBSHQSDSTFKDYATATYLDNONISSKSTNNPNK 773	

RESULT 2	
ABU01020	
ID	ABU01020 standard; protein; 2140 AA.
XX	
AC	ABU01020;
XX	
DT	23-OCT-2003 (revised)
DT	11-FEB-2003 (first entry)
XX	
DE	S. pneumoniae type 4 strain protein from coding region #590.
XX	
KN	Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection.

KM antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KM gene therapy; vaccine.

05 Streptococcus pneumoniae; type 4 strain.

PN WO200277021-A2.

03-OCT-2002

PF 27-MAR-2002; 2002WO-1B002163.

PR 27-MAR-2001; 2001GB-00007658.

PA (CHIR-) CHIRON SPA.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

XXIX

DR N-PSDB; ABX06302.

PT New proteins and

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or meningitis.

2
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

E

CC identity to an
CC modification

CC expressed from 2469 of 2489 identified
CC Streptococcus pneumoniae time 4 strain

AB565444. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus* nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to *Streptococcus* bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at ftp.wip0.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to standardise OS field)

Query Match	100.0%	Score 4026;	DB 6;	Length 2140;
Best Local Similarity	100.0%;	Pred. No. 2..9e-224;		
Matches 773; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

QY	LDGSIASKPKPNLNGKRGSLKKDTGYVHHHNNHNSIKKSSPTIRNISTIDFENK	60
Db	1334 KLGSIASKPKPNLNGKRGSLKKDTGYVHHHNNHNSIKKSSPTIRNISTIDFENK	139
QY	61 DLKTLKKKFRFVDPFTSETGKMEERYKRTDKNIIAYVDGTDLBYETEKLDISKSI	120
Db	1394 DLKGLIKKKFRFVDPFTSETGKMEERYKRTDKNIIAYVDGTDLBYETEKLDISKSI	145

```

QY 121 YGVLSPSKDGHEILGKISNVSKAKVYVYNNYKSIIEIKATKYDPSKMTEDLVANIND 180
DB 1454 YGVLSPSKDGHEILGKISNVSKAKVYVYNNYKSIIEIKATKYDPSKMTEDLVANIND 1513
QY 181 IVDGLAFAGDMRLPFVNDNDQKKAIRIKRPERIKETKSEFPYVSSYGAVYLEGEGDLSKN 240
DB 1514 IVDGLAFAGDMRLPFVNDNDQKKAIRIKRPERIKETKSEFPYVSSYGAVYLEGEGDLSKN 1573
QY 241 KPDNLTRESGKITYDSEKQOYLKDNITLARGYALKVTTPNGKTDMLRGVYSKEKI 300
DB 1574 KPDNLTRESGKITYDSEKQOYLKDNITLARGYALKVTTPNGKTDMLRGVYSKEKI 1633
QY 301 AKIQKAMPNLRALSETTIYADSRNVEDGSRSTQSVMSALDGNRIIRYQVTFPMNDKGA 360
DB 1634 AKIQKAMPNLRALSETTIYADSRNVEDGSRSTQSVMSALDGNRIIRYQVTFPMNDKGA 1693
QY 361 IDKGNLVTSSKLVLPFGDDEKXYTGDEKFNBAIKEDGSMFLPDTKPVNLSDMRNYPNP 420
DB 1694 IDKGNLVTSSKLVLPFGDDEKXYTGDEKFNBAIKEDGSMFLPDTKPVNLSDMRNYPNP 1753
QY 421 SKSNKTIYVRNPEFYLKRGKISDKGFGNMBLKVNESVVDNLTLYGDLHIDNTRDPNKLANK 480
DB 1754 SKSNKTIYVRNPEFYLKRGKISDKGFGNMBLKVNESVVDNLTLYGDLHIDNTRDPNKLANK 1813
QY 481 DGDIMDQKDYKANGFPDKVTDMDGNVYLQVSGDLNAKAVGVHYQFLYDNVKEBVND 540
DB 1814 DGDIMDQKDYKANGFPDKVTDMDGNVYLQVSGDLNAKAVGVHYQFLYDNVKEBVND 1873
QY 541 PKGNTSIEYADGKSVFVFNINDRRNNGFDEIOBOHIYNGKETYTFNDIKQIIDTLNLIK 600
DB 1874 PKGNTSIEYADGKSVFVFNINDRRNNGFDEIOBOHIYNGKETYTFNDIKQIIDTLNLIK 1933
QY 601 IVVNDPANTTYKCEPILANDGSEVSELKPHRYVTYITONGKEMSTTVSEEDTILPYKGE 660
DB 1934 IVVNDPANTTYKCEPILANDGSEVSELKPHRYVTYITONGKEMSTTVSEEDTILPYKGE 1993
QY 661 LEKGYQDFGWEISGFGKXKAGYVNLSPDPIKGVFKKIEBKKEBENKPTDVSKEKDN 720
DB 1994 LEKGYQDFGWEISGFGKXKAGYVNLSPDPIKGVFKKIEBKKEBENKPTDVSKEKDN 2053
QY 721 PQVNHSQLNESHKREDLQREHSQKSDSTKYTATVLDNNISSTSTNNPK 773
DB 2054 PQVNHSQLNESHKREDLQREHSQKSDSTKYTATVLDNNISSTSTNNPK 2106

```

RESULT 3

ABU45746
ID ABU45746 standard; protein; 2140 AA.

```

AC ABU45746;
DB 19-JUN-2003 (first entry)
DB Protein encoded by Prokaryotic essential gene #31273.
DB Antisense; prokaryotic essential gene; cell proliferation; drug design.
DB Streptococcus pneumoniae.
DB WO200277183-A2.
DB 03-OCT-2002.
DB 21-MAR-2002; 2002MO-US009107.
DB 21-MAR-2001; 2001US-00815242.
DB 06-SEP-2001; 2001US-00948993.
DB 25-OCT-2001; 2001US-0342923P.
DB 08-FEB-2002; 2002US-00072851.
DB 06-MAR-2002; 2002US-0362699P.
DB XX
DB PA (ELIT-) ELITRA PHARM INC.

```

```

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zvejkind JW,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
DB WPI; 2003-029926/02.
DB N-PSDB; ACA49616.
DB New antisense nucleic acids, useful for identifying proteins or screening
DB for homologous nucleic acids required for cellular proliferation to
DB isolate candidate molecules for rational drug discovery programs.
DB Claim 25; SEQ ID NO 73670; 1766bp; English.
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 623 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CC XX
SQ Sequence 2140 AA:
Query Match 100.0%; Score 4026; DB 6; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.9e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KGEIAESKFKMLGNGKESLKKDTTGVVHHQENBESITKESFTIDNISTIRDFENK 60
DB 1334 KGEIAESKFKMLGNGKESLKKDTTGVVHHQENBESITKESFTIDNISTIRDFENK 1393
QY 61 DLKKLIKKKFRVDDFTSETGRMEBYDYKDDKNIIVYDGTDLRYETKLDRIKSKI 120
DB 1394 DLKKLIKKKFRVDDFTSETGRMEBYDYKDDKNIIVYDGTDLRYETKLDRIKSKI 1453
QY 121 YGVLSPSKDGHEILGKISNVSKAKVYVYNNYKSIIEIKATKYDPSKMTEDLVANIND 180
DB 1454 YGVLSPSKDGHEILGKISNVSKAKVYVYNNYKSIIEIKATKYDPSKMTEDLVANIND 1513
QY 181 IVDGLAFAGDMRLPFVNDNDQKKAIRIKRPERIKETKSEFPYVSSYGAVYLEGEGDLSKN 240
DB 1514 IVDGLAFAGDMRLPFVNDNDQKKAIRIKRPERIKETKSEFPYVSSYGAVYLEGEGDLSKN 1573
QY 241 KPDNLTRESGKITYDSEKQOYLKDNITLARGYALKVTTPNGKTDMLRGVYSKEKI 300
DB 1574 KPDNLTRESGKITYDSEKQOYLKDNITLARGYALKVTTPNGKTDMLRGVYSKEKI 1633
QY 301 AKIQKAMPNLRALSETTIYADSRNVEDGSRSTQSVMSALDGNRIIRYQVTFPMNDKGA 360
DB 1634 AKIQKAMPNLRALSETTIYADSRNVEDGSRSTQSVMSALDGNRIIRYQVTFPMNDKGA 1693

```

QY 361 IDKDGMLVTDSSKLV.FGKDKKEYTGEDKFNVAIKEDSGMLPDTCPVNLSDMKRYFNP 420
 DB 1694 IDKDGMLVTDSSKLV.FGKDKKEYTGEDKFNVAIKEDSGMLPDTCPVNLSDMKRYFNP 1753
 QY 421 SKSNKIYVNPPEFYLAGKISDKGCFMWELRVNESVVDNYLIYGDHLIDNTRDFNFKLVK 480
 DB 1754 SKSNKIYVNPPEFYLAGKISDKGCFMWELRVNESVVDNYLIYGDHLIDNTRDFNFKLVK 1813
 QY 481 DGDINDMGKDYKANGPPDKVTDMDGNVTLQGTGYSDLNAAVGVHVOFLYDNVKEPVNID 540
 DB 1814 DGDINDMGKDYKANGPPDKVTDMDGNVTLQGTGYSDLNAAVGVHVOFLYDNVKEPVNID 1873
 QY 541 PKGNTSIEYADGKSVFENINDKRNNGPDEIOBOHIYINGKETSFPNDIKOIIDKTLNFK 600
 DB 1874 PKGNTSIEYADGKSVFENINDKRNNGPDEIOBOHIYINGKETSFPNDIKOIIDKTLNFK 1933
 QY 601 IYVKDPARTTVKEFIILNKDTGEVSELKPHRVTVTIQNGKEMSTTVSEEDFILPYKGE 660
 DB 1934 IYVKDPARTTVKEFIILNKDTGEVSELKPHRVTVTIQNGKEMSTTVSEEDFILPYKGE 1993
 QY 661 LEKGYPDGWEISGFGKGDAGVYNLSKDTPIKVPFKKIEKKEBENKPTFVDSKKDN 720
 DB 1994 LEKGYPDGWEISGFGKGDAGVYNLSKDTPIKVPFKKIEKKEBENKPTFVDSKKDN 2053
 QY 721 POWNHSQLNESHKEDLOREBHSOKSDSTKDVATATVLDKNNISSKSTNNPNK 773
 DB 2054 POWNHSQLNESHKEDLOREBHSOKSDSTKDVATATVLDKNNISSKSTNNPNK 2106

RESULT 4

ADM92113
 ID ADM92113 standard; protein; 2140 AA.

AC ADM92113;

DT 03-JUN-2004 (first entry)

DE S pneumoniae antigenic protein sequence SeqIDJ10.

KW antibacterial; gene therapy; Streptococcus pneumoniae infection;
 anti-genic.

OS Streptococcus pneumoniae.

PN WO2004020609-A2.

PD 11-MAR-2004.

PE 02-SEP-2003; 2003WO-US027401.

PR 30-AUG-2002; 2002US-0407082P.

RA (TUPF) UNIV TUPFS.

P1 Camilla A, Hava DL;

DR WPI; 2004-239189/22.

DR N-PSDB; ADM91876.

PT New Streptococcus pneumoniae nucleic acid molecules, useful for
 PT diagnosing, treating and preventing active infections of Streptococcus
 PT pneumoniae.

PS Claim 27; SEQ ID NO 310; 123pp; English.

CC This invention relates to novel isolated Streptococcus pneumoniae nucleic
 CC acid molecules and the antigenic polypeptides encoded by them. The
 CC invention may be useful for the production of compounds with an
 CC antibacterial activity or for gene therapy. The nucleic acid molecules,
 CC compositions and methods disclosed are useful for treating Streptococcus
 CC pneumoniae infection. The present sequence is that of an S pneumoniae
 CC protein of the invention.

XX SQ Sequence 2140 AA;

Query Match 100.0%; Score 4026; DB 8; Length 2140;
 Best Local Similarity 100.0%; Pred. No. 2,9e-224;
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLEFIASRYKNTANGKEGSLKQTTGVHEHHOENBSIYKSSPTIDRNIISTIRPEFK 60
 DB 1334 KLEFIASRYKNTANGKEGSLKQTTGVHEHHOENBSIYKSSPTIDRNIISTIRPEFK 1393
 QY 61 DLKGLIKKKREYDDFTSEFGKMEERYDYKYDDKGNIIAIDCGDLEFETKLDDEIKSI 120
 DB 1394 DLKGLIKKKREYDDFTSEFGKMEERYDYKYDDKGNIIAIDCGDLEFETKLDDEIKSI 1453
 QY 121 YGVLSPSKGFHFIILKISNVSNAAVYVYGNVYKSIIEIKATKYDFHSKWTFFDLYANIND 180
 DB 1454 YGVLSPSKGFHFIILKISNVSNAAVYVYGNVYKSIIEIKATKYDFHSKWTFFDLYANIND 1513
 QY 181 IYVGLAPAGDMRLFYVDNQKAEIILMPKIKETKSEVPYSSYGNVIELGEGDLSKN 240
 DB 1514 IYVGLAPAGDMRLFYVDNQKAEIILMPKIKETKSEVPYSSYGNVIELGEGDLSKN 1573
 QY 241 KPDMLTKEMSGKIYSDSEKQYLLKONIIILKGYALKVTTYNPKTMDLEGNGVYSKEDI 300
 DB 1574 KPDMLTKEMSGKIYSDSEKQYLLKONIIILKGYALKVTTYNPKTMDLEGNGVYSKEDI 1633
 QY 301 AKIQKANPMLRALSETTIYADSHNVEDGRTSGVLSALDGFNIIRYQVETFMNDKGA 360
 DB 1634 AKIQKANPMLRALSETTIYADSHNVEDGRTSGVLSALDGFNIIRYQVETFMNDKGA 1693
 QY 361 IDKDGMLVTDSSKLV.FGKDKKEYTGEDKFNVAIKEDSGMLPDTCPVNLSDMKRYFNP 420
 DB 1694 IDKDGMLVTDSSKLV.FGKDKKEYTGEDKFNVAIKEDSGMLPDTCPVNLSDMKRYFNP 1753
 QY 421 SKSNKIYVNPPEFYLAGKISDKGCFMWELRVNESVVDNYLIYGDHLIDNTRDFNFKLVK 480
 DB 1754 SKSNKIYVNPPEFYLAGKISDKGCFMWELRVNESVVDNYLIYGDHLIDNTRDFNFKLVK 1813
 QY 481 DGDINDMGKDYKANGPPDKVTDMDGNVTLQGTGYSDLNAAVGVHVOFLYDNVKEPVNID 540
 DB 1814 DGDINDMGKDYKANGPPDKVTDMDGNVTLQGTGYSDLNAAVGVHVOFLYDNVKEPVNID 1873
 QY 541 PKGNTSIEYADGKSVFENINDKRNNGPDEIOBOHIYINGKETSFPNDIKOIIDKTLNFK 600
 DB 1874 PKGNTSIEYADGKSVFENINDKRNNGPDEIOBOHIYINGKETSFPNDIKOIIDKTLNFK 1933
 QY 601 IYVKDPARTTVKEFIILNKDTGEVSELKPHRVTVTIQNGKEMSTTVSEEDFILPYKGE 660
 DB 1934 IYVKDPARTTVKEFIILNKDTGEVSELKPHRVTVTIQNGKEMSTTVSEEDFILPYKGE 1993
 QY 661 LEKGYPDGWEISGFGKGDAGVYNLSKDTPIKVPFKKIEKKEBENKPTFVDSKKDN 720
 DB 1994 LEKGYPDGWEISGFGKGDAGVYNLSKDTPIKVPFKKIEKKEBENKPTFVDSKKDN 2053
 QY 721 POWNHSQLNESHKEDLOREBHSOKSDSTKDVATATVLDKNNISSKSTNNPNK 773
 DB 2054 POWNHSQLNESHKEDLOREBHSOKSDSTKDVATATVLDKNNISSKSTNNPNK 2106

RESULT 5

AA81710
 ID AA81710 standard; protein; 2120 AA.

AC AA81710;

DT 02-JUN-2000 (first entry)

DE Streptococcus pneumoniae protein sequence IDJ.

KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
 bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
 kidney disease; diabetes; immunosuppressive disorder; otitis media;

KW pneumococcal septicemia; sinusitis; meningitis; therapy.
 OS Streptococcus pneumoniae.
 XX WO200006738-A2.
 XX 10-FEB-2000.
 XX 27-JUL-1999; 99WO-GB002452.
 XX 27-JUL-1998; 98GB-00016336.
 XX 19-MAR-1999; 99US-0125329P.
 XX (MICR-) MICROBIAL TECHNIQS LTD.
 XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
 XX WPI; 2000-195301/17.
 XX N-PSDB; AA291806.
 XX Streptococcal proteins and polynucleotides useful for diagnosis,
 XX treatment and prophylaxis of bacterial infections.
 XX Claim 2; Page 41-42; 76pp; English.
 XX This sequence represents a Streptococcus pneumoniae protein of the
 CC invention. The proteins (or their homologues, derivatives and/or
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 CC compositions comprising the proteins are useful as vaccines and also in
 CC diagnostic assays. The sequences are useful for the detection or
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
 CC with them. Agents capable of antagonising, inhibiting or interfering with
 CC the function or expression of the protein or polypeptide are useful in
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae
 CC infection. As the sequences can be used to treat S. pneumoniae infection,
 CC they can be used to treat bacterial pneumonia, which has high rates in
 CC young children, the elderly, and in patients with predisposing conditions
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 CC or with immunosuppressive disorders, especially AIDS. They can also be
 CC used to treat pneumococcal septicemia, otitis media, sinusitis, and
 CC meningitis
 XX Sequence 2120 AA;
 SO
 Query Match 99.7%; Score 4015.5; DB 3; Length 2120;
 Best Local Similarity 99.9%; Pred. No. 1.1e-223;
 Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 KLGGIABSKFKNLGKSGSLKADDTTGVHHHQQENBSIKKSSPTIDNISTIRDPENK 60
 DB 1313 KLGGIABSKFKNLGKSGSLKADDTTGVHHHQQENBSIKKSSPTIDNISTIRDPENK 1372
 QY 61 DLKKLIKKKFRVDDPTSETGRMEYDYKYDDKGNIIAYDDGTDLLEYETKLDIKSKKI 120
 DB 1373 DLKKLIKKKFRVDDPTSETGRMEYDYKYDDKGNIIAYDDGTDLLEYETKLDIKSKKI 1432
 QY 121 YGVLSPPSKDGFHEIIGKISNVSNAKAVYVGNYSIIRATYDFPSKMTFDLYANIND 180
 DB 1433 YGVLSPPSKDGFHEIIGKISNVSNAKAVYVGNYSIIRATYDFPSKMTFDLYANIND 1492
 QY 181 IVDGLAPAGDMPLPTKNDQKAAETKIRPBEIKETKSPYVSSGVNVIIRGEGDLSN 240
 DB 1493 IVDGLAPAGDMPLPTKNDQKAAETKIRPBEIKETKSPYVSSGVNVIIRGEGDLSN 1552
 QY 241 KPDNLTKKSGKCIYDSEKQYTLKDNITLRGVALKVTYVPGKIDMLGNGVYSKEDI 300
 DB 1553 KPDNLTKKSGKCIYDSEKQYTLKDNITLRGVALKVTYVPGKIDMLGNGVYSKEDI 1612
 QY 301 AKIQANRYLRALSTETIYADSRNVEDGSGTOSVLSALDGNRIIRYQVPTFRANDKGA 360
 DB 1613 AKIQANRYLRALSTETIYADSRNVEDGSGTOSVLSALDGNRIIRYQVPTFRANDKGA 1672
 QY 361 IDKQGNLVTDSSKLVLFQKDDREYTGEDKFNVEAIKEDSMLFIDTKPVLNMDKNYFNP 420

DB 1673 IDKQGNLVTDSSKLVLFQKDDREYTGEDKFNVEAIKEDSMLFIDTKPVLNMDKNYFNP 1732
 QY 421 SKSNKTYVNPPEFYLAGKISDGGFWMLRYVESVVDNLTLYGDLHIDNFR-DPNIKLV 479
 DB 1733 SKSNKTYVNPPEFYLAGKISDGGFWMLRYVESVVDNLTLYGDLHIDNFRDPNISKLV 1792
 QY 480 KQGDINDMGKDYKANGPPDKVTMDGNVYLQGYSDLNAKAVGHYQFLYDNVCKEVI 539
 DB 1793 KQGDINDMGKDYKANGPPDKVTMDGNVYLQGYSDLNAKAVGHYQFLYDNVCKEVI 1852
 QY 540 DPKGNTSIRYADGKSVYFNINDKRNNGFGEIOEGHIIYNGKETSFPNDIKOIIDKTLNI 599
 DB 1853 DPKGNTSIRYADGKSVYFNINDKRNNGFGEIOEGHIIYNGKETSFPNDIKOIIDKTLNI 1912
 QY 600 KIVVQDFARTWYKKEITLKKDGEVSELKPHRTVYTIQNGKMSPTVSEEDILPYK 659
 DB 1913 KIVVQDFARTWYKKEITLKKDGEVSELKPHRTVYTIQNGKMSPTVSEEDILPYK 1972
 QY 660 ELEKGYPDQWBIISGFEGKKDAGVYVNLKDTFIEKVFCKI EKKKEENKPTDVSKKD 719
 DB 1973 ELEKGYPDQWBIISGFEGKKDAGVYVNLKDTFIEKVFCKI EKKKEENKPTDVSKKD 2032
 QY 720 NPQVNSQANESHKREDLQREHSQKSDSTKVYATVLDKNNISSTGTNNPK 773
 DB 2033 NPQVNSQANESHKREDLQREHSQKSDSTKVYATVLDKNNISSTGTNNPK 2086
 RESULT 6
 ADK48759
 ID ADK48759 standard; protein; 2138 AA.
 XX ADK48759;
 AC 20-MAY-2004 (first entry)
 XX 20-MAY-2004 (first entry)
 DT Streptococcus pneumoniae protein, Seq ID No 5274.
 XX Streptococcus pneumoniae protein, Seq ID No 5274.
 XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 XX US6699703-B1.
 PN 02-MAR-2004.
 PD 02-MAR-2004.
 XX 26-MAY-2000; 2000US-00583110.
 PF 02-JUL-1997; 97US-0051553P.
 PR 12-MAY-1998; 98US-0085131P.
 XX 30-JUN-1998; 98US-00107433.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PA Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CB;
 PI N-PSDB; ADK46098.
 DR N-PSDB; ADK46098.
 XX New nucleic acid molecules and polypeptides useful for diagnosing,
 PT preventing and treating pathological conditions resulting from bacterial
 PT infection, e.g. Streptococcus pneumoniae infection, and in drug
 PT screening.
 XX Disclosure; SEQ ID NO 5274; 301bp; English.
 PS The invention relates to isolated Streptococcus pneumoniae nucleic acids
 XX and polypeptides. The nucleic acids and proteins are useful for
 CC diagnosing, preventing and treating pathological conditions resulting
 CC from bacterial infection, such as S. pneumoniae infection. These may also
 CC be used for drug screening procedures. The present sequence represents a
 CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
 CC data for this patent did not appear in the printed specification but was

CC obtained in electronic format directly from USPTO at
 CC seqdata.uspto.gov/sequence.html.
 XX
 SO Sequence 2138 AA;

Query Match 94.1%; Score 3789.5; DB 8; Length 2138;
 Best Local Similarity 94.8%; Pred. No. 1.5e-210;
 Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;

```

QY 1 KGEIASEKRNKNGKESLKKDTTGVHHHQNESISEKSFPTDRIINSTRDENEK 60
DB 1333 KGEIASEKRNKNGKESLKKDTTGVHHHQNESISEKSFPTDRIINSTRDENEK 1392
QY 61 DLKCLIKKKRREVDPTSETGKMESEYDYKDDKGNIAVDGDTLETEKLEDEIKSI 120
DB 1393 DLKCLIKKKRREVDPTSETGKMESEYDYKDDKGNIAVDGDTLETEKLEDEIKSI 1451
QY 121 YGVLSKSGDGHFILLKISNVSKAKAYYGNNTKSHIKATKDFHSKWTFFDYANIND 180
DB 1452 YGVLSKSGDGHFILLKISNVSKAKAYYGNNTKSHIKATKDFHSKWTFFDYANIND 1511
QY 181 YVGLAFAGDMRLFVKNDDOKKARIKIRMEPKIKETSEKPYSSYGNVIELSGDLSKN 240
DB 1512 YVGLAFAGDMRLFVKNDDOKKARIKIRMEPKIKETSEKPYSSYGNVIELSGDLSKN 1571
QY 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIIRKGYALKVTTYNGKTDMLGNGVYSKEDI 300
DB 1572 KPDNLTKMESGKIYSDSEKQYLLKDNIIIRKGYALKVTTYNGKTDMLGNGVYSKEDI 1631
QY 301 AKIQKRNPLRALSETTIVADSRNVEDSGTOSVLSALDGFNIIIRYQVTFPKNDRKGA 360
DB 1632 AKIQKRNPLRALSETTIVADSRNVEDSGTOSVLSALDGFNIIIRYQVTFPKNDRKGA 1691
QY 361 IDKQNLVYDSSKLVLFQKDDKEYTGBDKFNVAIKEDGSMFLFDTPVNLSDMKYFNP 420
DB 1692 IDKQNLVYDSSKLVLFQKDDKEYTGBDKFNVAIKEDGSMFLFDTPVNLSDMKYFNP 1751
QY 421 SSKNKTIVNPEPTLKGKISDKGFPNWEIRVNESVDNNTLIYGDLDHIDTRDNIKLNVK 480
DB 1752 SSKNKTIVNPEPTLKGKISDKGFPNWEIRVNESVDNNTLIYGDLDHIDTRDNIKLNVK 1811
QY 481 DDGIDMGKDDKANKANPPKVTMDGNVYTCYGSPLNAAVGVHOFLYDNKRPENID 540
DB 1812 DDGIDMGKDDKANKANPPKVTMDGNVYTCYGSPLNAAVGVHOFLYDNKRPENID 1871
QY 541 PKGNTSIRBYADGKSVFENIDKRNNGPDCGELIQOHLYINGKEYTSFNDIKQIIDKTLNIX 600
DB 1872 PKGNTSIRBYADGKSVFENIDKRNNGPDCGELIQOHLYINGKEYTSFNDIKQIIDKTLNIX 1931
QY 601 IYVKOPARNTTYKEPILNIDTGEVSLKPHRVTVTIUNGKEMSGTIVSEEDFILPYKGE 660
DB 1932 IYVKOPARNTTYKEPILNIDTGEVSLKPHRVTVTIUNGKEMSGTIVSEEDFILPYKGE 1991
QY 661 LEKGYOPDGEWISGFEKGDAGVYINLSKDTPIKPPKIEKKECEBENPTFVYSKKN 720
DB 1992 LEKGYOPDGEWISGFEKGDAGVYINLSKDTPIKPPKIEKKECEBENPTFVYSKKN 2051
QY 721 POWNHSHLNESHKEDLQREBHSQKSDSTKYATVATVLDKNNISKSSTNNPNK 773
DB 2052 POWNHSHLNESHKEDLQREBHSQKSDSTKYATVATVLDKNNISKSSTNNPNK 2104

```

RESULT 7
 ID ADR94534 standard; protein; 637 AA.
 XX ADR94534;
 XX AC
 XX ADR94534;
 DT 16-DEC-2004 (first entry)
 XX
 XX Novel S. pneumoniae protein sequence, SEQ ID 3169.
 XX
 KW Meningitis; bacteriaemia; pneumonia; otitis media; vaccine;

```

KW bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
EN US6800744-B1.
XX
PD 05-OCT-2004.
XX
PF 30-JUN-1998; 98US-00107433.
XX
PR 02-JUL-1997; 97US-0051553P.
XX
PR 12-MAY-1998; 98US-0085131P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2004-697205/68.
XX
DR N-PSDB; ADR91931.
XX
XX
PT New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.
XX
PS Disclosure; SEQ ID NO 3169; 151pp; English.
XX
XX
CC The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91166polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridizable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
XX
SO Sequence 637 AA;

```

Query Match 75.7%; Score 3048; DB 8; Length 637;
 Best Local Similarity 99.3%; Pred. No. 2.8e-168;
 Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

QY 185 LAFAGDMRLFVKNDDOKKARIKIRMEPKIKETSEKPYSSYGNVIELSGDLSKNPND 244
DB 15 LAFAGDMRLFVKNDDOKKARIKIRMEPKIKETSEKPYSSYGNVIELSGDLSKNPND 74
QY 245 LTRMESGKIYSDSEKQYLLKDNIIIRKGYALKVTTYNGKTDMLGNGVYSKEDI 304
DB 75 LTRMESGKIYSDSEKQYLLKDNIIIRKGYALKVTTYNGKTDMLGNGVYSKEDI 134
QY 305 KANPNLRALSETTIVADSRNVEDSGTOSVLSALDGFNIIIRYQVTFPKNDRKGAIDND 364
DB 135 KANPNLRALSETTIVADSRNVEDSGTOSVLSALDGFNIIIRYQVTFPKNDRKGAIDND 194
QY 365 GNLVYDSSKLVLFQKDDKEYTGBDKFNVAIKEDGSMFLFDTPVNLSDMKYFNPFSKN 424
DB 195 GNLVYDSSKLVLFQKDDKEYTGBDKFNVAIKEDGSMFLFDTPVNLSDMKYFNPFSKN 254
QY 425 KIYVNPPEPTLKGKISDKGFPNWEIRVNESVDNNTLIYGDLDHIDTRDNIKLNVKXGDI 484

```

```

DB 255 KTYANPFTYLGKISDKGSFNNWELAVNESVDNVTITGDHLDNTRDNRIKLANVKGDI 314
QY 485 KDWGKDYKANGFPDKVTMDGANNVYQTSYSDLANAKAVGVRQFLYDNYKPEVNIIDPKCN 544
DB 315 KDWGKDYKANGFPDKVTMDGANNVYQTSYSDLANAKAVGVRQFLYDNYKPEVNIIDPKCN 374
QY 545 TSIEVADKSVFVFNDRKNNNGPDGEIOGHYINKEKYSFNDIQIIDKTNITIVVK 604
DB 375 TSIEVADKSVFVFNDRKNNNGPDGEIOGHYINKEKYSFNDIQIIDKTNITIVVK 434
QY 605 DPARVTVYKFTLNKDTGSEVSLKPRVYVTTIQNGKMSSTIYSEEDFLPYKGLKLG 664
DB 435 DPARVTVYKFTLNKDTGSEVSLKPRVYVTTIQNGKMSSTIYSEEDFLPYKGLKLG 494
QY 665 YQPDGMEISGFEKCKDAGTVINLSKDTPIKPVPKLIEKKEENKPTPVSKKXDPQVN 724
DB 495 YQPDGMEISGFEKCKDAGTVINLSKDTPIKPVPKLIEKKEENKPTPVSKKXDPQVN 554
QY 725 HSQLESRKEDLQREBSQKSDSTQVATVLDKNISKSTNNPNK 773
DB 555 HSQLESRKEDLQREBSQKSDSTQVATVLDKNISKSTNNPNK 603

```

RESULT 8
ADP96136
ID ADP96136 standard; protein; 1529 AA.

ADP96136;
16-DEC-2004 (first entry)

Novel S. pneumoniae protein sequence, SEQ ID 4771.

Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
bacterial infection.

Streptococcus pneumoniae.

US600744-B1.

05-OCT-2004.

30-JUN-1998; 98US-00107433.

02-JUN-1997; 97US-0051553P.

12-MAY-1998; 98US-0085131P.

(GENO-) GENOME THERAPEUTICS CORP.

Doucette-Stamm LA, Bush D;

WPI; 2004-697205/68.

N-PSDB; ADP93533.

New isolated nucleic acid encoding a Streptococcus pneumoniae
polypeptide, useful for diagnosing, preventing and/or treating
pathological conditions resulting from the bacterial infection.

Disclosure: SEQ ID NO 4771; 151pp; English.

The invention relates to an isolated nucleic acid comprising a sequence
encoding a Streptococcus pneumoniae ADP91366polypeptide, or its
fragments, with any of 9 fully defined sequences (appearing as ADP94308,
ADP94489, ADP94800, ADP94837, ADP94969, ADP95253, ADP95642, ADP95682,
ADP96079) or any of the fully defined sequences appearing as ADP91705,
ADP91886, ADP92197, ADP92234, ADP93039, ADP93079, ADP92366, ADP92650 or
ADP93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
hybridisable under high stringency conditions to the nucleotide sequence.
The nucleic acid and proteins are chosen from 5206 disclosed sequences.
Also included are a recombinant expression vector comprising the isolated
nucleic acid cited above operably linked to a transcription regulatory

CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.

Sequence 1529 AA;

Query Match 18.6%; Score 750.5; DB 8; Length 1529;
Best local Similarity 79.7%; Pred. No. 1.3e-34;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

```

QY 1 KLGRIASRKNLGNKGESLKKDTGVEHHHQENESIKESKSPITDRNISTIRDFENK 60
DB 1336 KLGRIEBSKFNKLVKXKDSLNKRTAVENNLVYDQSLGKSLFVHHTKISTIRDFENK 1395
QY 61 DLKGLIKKKFREVDFTSFTGKRMERYDYKYDDKNIIVYDQDTLREYETKLDKISKI 120
DB 1396 DLKGLIKKKYKQEDDFVN-GGTRVARDYKYDDKNIIVYDQDTLREYETKLDKISKI 1454
QY 121 YGVLSRKHGPEILAKISNVSKNAKVYGGNNYSIETATKYDFHSKMTFPLVYANIND 180
DB 1455 YGVLSRKHGPEILAKISNVSKNAKVYGGNNYSIETATKYDFHSKMTFPLVYANIND 1514
QY 181 IVDGLAF 187
DB 1515 IVDGISF 1521

```

RESULT 9

AAW55096
ID AAW55096 standard; protein; 117 AA.

AAW55096;

02-OCT-1998 (first entry)

Streptococcus pneumoniae SP0043 protein.

Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
detection; pneumonia; otitis media; meningitis.

Streptococcus pneumoniae.

WO9818930-A2.

07-MAY-1998.

30-OCT-1997; 97WO-US019422.

31-OCT-1996; 96US-0023960P.

(HUMA-) HUMAN GENOME SCI INC.

Kunach CA, Choi GH, Johnson LS, Hromocky J A;

WPI; 1998-272224/24.

N-PSDB; AAW7357.

Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
- or their epitope-containing fragments, useful in protective or
therapeutic vaccines, and for diagnosis.

Claim 11; Page 62; 118pp; English.

The present sequence represents a protein from Streptococcus pneumoniae.
The nucleic acid sequence encoding the Streptococcus pneumoniae protein

CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose

CC pneumoniae antigens have antibacterial activity and can be used in
 CC vaccines. The S. pneumoniae antigens can also be used to prevent or
 CC attenuate a Streptococcal infection in an animal. The polynucleotides
 CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
 CC nucleic acids. AB084905 to AB085130 represent primers used in the cloning
 CC of S. pneumoniae ORFs (open reading frames) which are used in an example
 CC from the present invention

XX Sequence 117 AA:

Query Match 15.3%; Score 615; DB 2; Length 117;
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGMEISGFGKKGADGVYINLSKDTFIKPVFKIEKKEENKPTDVSK 716
 DB 1 YKGELEKGYQPDGMEISGFGKKGADGVYINLSKDTFIKPVFKIEKKEENKPTDVSK 60

QY 717 KKNPQVNHSQLNESHKEDLQREHRSQKSDSTKDYATATVLDKNNTSSKSTNNPK 773
 DB 61 KKNPQVNHSQLNESHKEDLQREHRSQKSDSTKDYATATVLDKNNTSSKSTNNPK 117

RESULT 10
 ABP54590
 ID ABP54590 standard; protein; 117 AA.

XX ABP54590;

DT 04-SEP-2002 (first entry)

XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

OS US2002061545-A1.

XX 23-MAY-2002.

XX 22-JAN-2001; 2001US-00765272.

XX 30-OCT-1997; 97US-00961083.

XX (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.

PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;

DR WPI; 2002-479261/51.
 DR N-PSDB; AB084825.

PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
 PT and for preventing or attenuating disease caused by Streptococcus
 PT infection.

PS Claim 11; Page 29; 70pp; English.

CC AB084792 to AB084904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.

XX Sequence 117 AA:

Query Match 15.3%; Score 615; DB 5; Length 117;
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQPDGMEISGFGKKGADGVYINLSKDTFIKPVFKIEKKEENKPTDVSK 716
 DB 1 YKGELEKGYQPDGMEISGFGKKGADGVYINLSKDTFIKPVFKIEKKEENKPTDVSK 60

QY 717 KKNPQVNHSQLNESHKEDLQREHRSQKSDSTKDYATATVLDKNNTSSKSTNNPK 773
 DB 61 KKNPQVNHSQLNESHKEDLQREHRSQKSDSTKDYATATVLDKNNTSSKSTNNPK 117

ADCA5149

XX ADCA5149;

DT 18-DEC-2003 (first entry)

XX S. pneumoniae antigenic protein SP043.

XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.

XX Streptococcus pneumoniae.

OS US6573082-B1.

XX 28-MAR-2000; 2000US-00536784.

XX 31-OCT-1996; 96US-0029960P.

XX 30-OCT-1997; 97US-00961083.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
 PI Rosen CA;

DR WPI; 2003-764574/72.
 DR N-PSDB; ADCA5148.

PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
 PT useful for producing vaccines for prevention or attenuation of infection
 PT by Streptococcus pneumoniae.

PS Example 1; SEQ ID NO 68; 58pp; English.

XX The invention relates to an isolated polynucleotide consisting of a
 CC Streptococcus pneumoniae nucleic acid (appearing as ADCA5122 and encoding
 CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
 CC antigens. Also included are making a recombinant vector by inserting the
 CC nucleic acid into a vector, an isolated polynucleotide consisting of at
 CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
 CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
 CC acids are useful as DNA vaccine against Streptococcus pneumoniae
 CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
 CC antigen nucleic acids are useful as probes for use in diagnostic methods
 CC for detecting S. pneumoniae gene expression. The present sequence
 CC represents an S. pneumoniae antigenic protein.

SQ Sequence 117 AA;
 Query Match 15.3%; Score 615; DB 7; Length 117;
 Best Local Similarity 100.0%; Pred. No. 3.2e-28;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 657 YKGELEKGYOPDGEISGFECKDAGYVNLKDTPIKPVKIKI EKKKEENKKTPTVSK 716
 DB 1 YKGELEKGYOPDGEISGFECKDAGYVNLKDTPIKPVKIKI EKKKEENKKTPTVSK 60
 QY 717 KCONPVNHSQNLNESHKEDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 773
 DB 61 KCONPVNHSQNLNESHKEDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 117
 RESULT 12
 ABU24813
 ID ABU24813 standard; protein; 1196 AA.
 XX ABU24813;
 AC
 XX 19-JUN-2003 (first entry)
 DT
 XX Protein encoded by Prokaryotic essential gene #10340.
 DB
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KM
 XX Clostridium botulinum.
 OS
 XX WO200277183-A2.
 PN
 XX 03-OCT-2002.
 PD
 XX 21-MAR-2002; 2002MO-US009107.
 XX
 PF 21-MAR-2001; 2001US-00815242.
 XX
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362599P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyekind JW,
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu RH;
 PI WPI: 2003-029926/02.
 DR N-BSDB; ACA28683.
 DR
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 PT
 XX Claim 25; SEQ ID NO 52737; 1766bp; English.
 PS
 XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the antisense
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 SQ Sequence 1196 AA;
 Query Match 5.6%; Score 227; DB 6; Length 1196;
 Best Local Similarity 20.2%; Pred. No. 0.00021;
 Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;
 QY 1 KLGRIASKEKFNGLNGKRG-----SLKKDTTGVHHEHNESESI 39
 DB 123 RLKDIOE-LFMDVGIGKGYSIIGQKIEAVLSGKPEERSLLEAGIYFKRKXBEAD 181
 QY 40 KEKSSFTIDNISTIPDEN--KDKKLICKTREVDF--TSETGRME----- 85
 DB 182 KKLNS--TEQNLIRKDIILATYERMEPLKEBEKAKPFLNLSBLRKEENVNVIYSIDK 239
 QY 86 -EYDYKYDDKGNIIAVIDDGT-LVEYETEKDEIKSKYGLVLSKXGHPFELIGISVSK 143
 DB 240 IERDILK-NISSNLSIGENDLNKREKSYKKEIISK-----NEKILYLDK--NNSR 288
 QY 144 NARVYGNANNYSIEIKATKYDFHSKTFPDLVANINDYDGLAPAGDMRLFVNDOPKA 203
 DB 289 NKEEYYNKKDKNDIR-----NENALTKERIKLKD-----NIKVENLTAKTN 331
 QY 204 EIKIR-----MPEKIKETSEYPPV----- 223
 DB 332 EERLQYLYKEKALEKKKINLKEBESYLSIEIGKEDNANNFNKLEKEKCLKILKSEB 391
 QY 224 -----SSYGVNIEIGEDL--SKNPDNLTPWESGKYS-----DSEKQYLLK 265
 DB 392 IELNSTSMRLNBSISWENBIYVLEKNLKNINRSCSYISSININIKTKEDIEKINIK 451
 QY 266 DNITL-----KRGVALKTTVPKGTMLEGNGVYSKED-----IAKIOK--AN 307
 DB 452 ENITLLENNLKENSRNISLKIISLNKKEK-KLKEKNAAYSRIBANYHMLNLEKHYEGYN 510
 QY 308 PNLRALSETTIYADSNVDESGTOSVMSALDGFNIIRYQVTFPKNKGKGAIDDGNL 367
 DB 511 RSVKTLMEVSYGKVDNINIG-----CEVLGDIIIKYKKELETAMEIALGGA1--SNV 560
 QY 368 VTDSSKLVLGKDDKEYTEGDEKFNVEAIXEDSGMLPDPKPVLSMDKXVFNPD--SKSN 424
 DB 561 ITEDS-----NKAKILNLYLKKK-----SLGRATFLPLTTIGR 594
 QY 425 KLYVRPEYTLGAKISDKGPFWEALRVNESVVDNYLYGDLIINDTRDFT--KLNVKDG 492
 DB 595 KAKINN-----VIREDF--LGIASDLIDY-----DVRSNIIDVYLGRTLLAKD- 637
 QY 483 DIMDWMKDYKANGPFDKTDMDGNVYLGTVSDLNAAKVGHYGLYNVNEPVNIDPK 542
 DB 638 --MDSALKIAGLANYSFKIVTLEGV-----INFG 665
 QY 543 GNT--SIERYADKSVFENINIKRNNGFDGEIIOGHIIYNGKESYSPNDIKOIIDTTLNI 599
 DB 666 GSLTGGISIGIRAGSSI--ISRR-----RIER-----TKLEBETKN- 700
 QY 600 KIVVDFANNTVKEPIFNKONGEVSLEKPHRVYTTIQNGKEMSSYVEBPFLPYVYG 659
 DB 701 -----TIEFMEK-----ILEKNKIKT--LDEENLNI--KD 728
 QY 660 BLEKGYOPDGEISGFECKDA-----GVVINLSKD--TFIKPVFKIKI EKKKEENK 709

PR 27-AUG-2002; 2002US-0406676P.
 PR 28-AUG-2002; 2002US-0406757P.
 PA (CHTR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Tettein H, Masjmani V;
 DR MPI, 2004-248071/23.
 XX
 PT Immunogenic composition useful as a vaccine for treating or preventing
 PT streptococcal infections, comprises group B Streptococcus polypeptides.
 XX
 PS Claim 10; SEQ ID NO 4422; 1194pp; English.
 CC
 CC The invention relates to a novel immunogenic composition comprising a
 CC combination of 2-5 group B Streptococcus (GBS) polypeptides. Each
 CC polypeptide is encoded by a GBS polynucleotide sequence which is
 CC homologous to a polynucleotide sequence of group A Streptococcus (GAS),
 CC Streptococcus pneumoniae and/or at least one other GBS serotype. The
 CC composition of the invention demonstrates antibacterial activity whilst
 CC the polypeptides and polynucleotides may be useful in assays to diagnose
 CC and identify streptococcal infections or for identifying, screening and
 CC developing vaccines and other treatments for streptococcal infections.
 CC The current sequence is that of a Streptococcus agalactiae ORF SAG
 CC protein of the invention.
 CC
 SQ Sequence 1166 AA;

Query Match 5.5%; Score 220; DB 8; Length 1166;
 Best Local Similarity 20.2%; Pred. No. 0.00051;
 Matches 198; Conservative 139; Mismatches 306; Indels 336; Gaps 56;

QY 4 EIAESKPKNNKANGKSGSLKQTTGVHHHQBESIKKSSFTIDNISTINDFENKDLK 63
 DB 60 EIEBELSKNLDTSNMGA-----DLSEYPSKPEITNNYKS-NVVTNASTIA-----IA 105
 QY 64 KLIKKEKPREVDFTS-----ETGKMEBYDYKDDKGNIA-YDDGDLSEYTEKLDDEI 116
 DB 106 QKVPAYAEVKKKSSSLAVLDTSKITKQAITQKRGKVVAVLIDGPFINDHFDLD-- 163
 QY 117 KSKITGVLSPSKQGH-----PELIGKISNVS---KNAKYVYGNVYS---TEIYA 160
 DB 164 -----SPKDDKISFKKAKAFEBELKAKNITVYGMVNDKIVFANVANNETVADIA 215
 QY 161 TKYDHSKMTFTDVLVANDIDVGLAFAGDMELFYVDNOKKA-----EIK 206
 DB 216 AMKDSIG-----SEANNISHGTHVAG--TFV-GNSKRPAINGLLEGAAPNAQVL 263
 QY 207 IRMPKIKETKSEYPVSSYGVNVIIEGDLSKNKPDLTKMESGR-----IYDSERKQY 262
 DB 264 MRIPKIDSKFGBAVAKAITDAVNLGATIN-----MSIGTASLIALANDKVL 314
 QY 263 LKQDNIIILKGYALVYTTNPGKTMLBGNVYSKEDIKIQGANP-----NIRALSETT 317
 DB 315 ALK--LASERKGVAVVVAAGNBAFGM-----DYSEKPLSTNDPYGVNSPAISEDT 362
 QY 318 IYADSRNVEDGRSTQSVLMSALDGFNIIRYGVTFPGMNRKBAIDKQMLVYDSSGLVLF 377
 DB 363 LSVAS--YESLKTISEVETITIEG-KLVKLPVITSFPPKRGKAYD-----VVY 407
 QY 378 GK--DDKRYTGD--KENVAIKEDGSMFLID--TKFVNLSMDKQVFNPSKNTIYRANB 432
 DB 408 ANYGAKDFEGDFKQKIALIERGGGLDPMTKITHATNGV-----VGIVAFENDQ 457
 QY 433 FYLRKSIDKXGPFNMEIARVNESVDNYLL-YGDL-----HIDNTRD--FNITL 477
 DB 458 -----EKRG-----NPLPYRELPGVATSKYDGERIKNTISQGLTGNQRP 496
 QY 478 NVKD--GDIM-----DWGM-----KDYKANGFPDKVTDMDQNVVLTQGYSDINAKAVG 523
 DB 497 EYVDGQGNRMLEQSSWSVTAAGAIKPDVTAGSEFIYSSTYNNQVOTMGTSMASHAVAG 556

QY 524 V-----HYQFLVDNYKPEVNIIDPK-----GNTSIEYADGKSVFVNIDKRNNG 566
 DB 557 LMTMLQSHLAEKRG-----MNLDSKGLLEBSKNILMSASATALYSBBDKAFYSRQAGV 612
 QY 567 FDGE--IQGQHIYNGKERTSFNDIKQIITDKTANIKIYVKDPFARNTTKEFPIINKDTGEV 624
 DB 613 VDAEKAIQAOY-VYVNDGKAKINLRVGDK-FDITVTIHKLVGEVVELYQANVATEQV 670
 QY 625 SE--LKPH-----RVTVTIQ-----NG-----639
 DB 671 NKGKPAKFAQALDITWQVILADKETQVRFITIDASQFSQKLKEQVANGYFLEGVPRKE 730
 QY 640 -KEMSSIVS-----EBDFI-----LPYKGELEKGY-----QPDGWEISG 674
 DB 731 AKDSNQELMSTIPVGFNPGFASLQALEPIYK-TLSKSPFYVKNPTTHDQLEYNESAP 789
 QY 675 FEGSK-----DAGYV-----INLSKDT--FKPVPK-KIEEK-----KEES 707
 DB 790 FESNNYTAALLTQSASGSDYVYKNGGELIAPESPRIILIGPENKVEDTTHILBRDA 849
 QY 708 NKPTFVSXKXDPQVNHSQLNBSHREKEDLOREHSQSD--STRQVATVLDKN-NI- 762
 DB 850 NMPYFAISPKDG-----NRDEITPQATFLANVKDISAQVLDQNGVNI 892
 QY 763 -----SKSTNNPNK 773
 DB 893 MQSKVLPSYRKQNFNNPKQ 911

RESULT 15
 AAM54145
 ID AAM54145 standard; protein; 1639 AA.
 XX
 AC AAM54145;
 XX
 DT 25-MAR-2003 (revised)
 DT 23-SEP-1998 (first entry)
 XX
 DE P. falciparum synthetic gp190 protein.
 XX
 KW gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
 KW monoclonal antibody; passive immunisation; parasite.
 OS Plasmodium falciparum.
 OS Synthetic.
 PN MO9814583-A2.
 XX
 PD 09-APR-1998.
 XX
 PF 02-OCT-1997; 97MO-BP005441.
 XX
 PR 02-OCT-1996; 96DE-01040817.
 XX
 PA (BUQA/) BUJARD H.
 XX
 PI Bujard H, Tolle R, Pan W;
 XX
 DR MPI, 1998-240088/21.
 DR N-PSDB; AAV21451, AAV35363.
 XX
 PT Recombinant production of complete gp190/MSP-1 Plasmodium surface protein
 PT - useful in anti-malaria vaccines, also stabilising genes by reducing
 PT their AT content.
 XX
 PS Example 1; Fig 3c; 48pp; German.
 CC This sequence represents a modified Plasmodium falciparum gp190/MSP-1
 CC (merozoite surface) protein. The gene encoding this protein has been
 CC stabilised by reducing the AT content of the nucleotide sequence. Such a
 CC protein is useful in vaccines against malaria or for producing monoclonal
 CC antibodies (for passive immunisation). The complete gp190 protein can now
 CC be produced outside the parasite and has, at least over extended regions,

CC the native pattern of folding. Larger amounts of the protein can be
 CC produced recombinantly than would be possible using the parasites as
 CC source. (Updated on 25-MAR-2003 to correct PR field.)

XX
 SQ Sequence 1639 AA;

Query Match 5.4%; Score 218; DB 2; Length 1639;

Best Local Similarity 18.6%; Pred. No. 0.0011;
 Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

```

QY 23 KQTG-VEHHQNESIK-----EKSFTIDRNTSTIDFENKDLKKKKFREVD 75
DB 232 KQVGMEDYIKKKKTIENINELIBSKTIDKKNVATKBER-----KKLYQAQY 283
QY 76 FTSFGKREEDYKATDCKGNIAYDGTDLAYETBKLDRIKSKIYGLVSPKDGFEIL 135
DB 284 DLSIYKQLEB-----AHNLISV-----LB---KRIIDTLK-----KQENIKEL 319
QY 136 GKISNVSKAKVYVGNYSIKIKATKYDPH-----SKTMTFDLYANINDIVDGLAFA 188
DB 320 DKINBIKNPPANGTPTNLTLDKNNKKIBHBKIKIATIKFNIDSLFTDPL----- 373
QY 189 GMRRLPVKNDOKAKIKIRMPKIKETSEY---YSSYGV---TELGE-GLSK 239
DB 374 -ELEYYLRBN- KNIDISAKVETKSTEPPEYNGVYTPLSYNDINNALNELNSFGDLI- 430
QY 240 NRPDNLTKRESGKIYDSERKOYL--LKNHIL-----RKGVALKVTYVNGKTDML 289
DB 431 -NPDYTKRPSKNITVDNERKIKINIKIKIKIKIKESDKSYEDRSKSLN---DIT 485
QY 290 EG---NGVYSKE---DIAKIQRA---NPNRLASETTYADSRNVEDGSTOSV 334
DB 486 KEYEKLIINIDSKFNNNIDLTNFERKMGKRYSVKYEKLTHTNTPASYENSKNLIBKLT 545
QY 335 LMSALDGF---NI-----RYGVTFPKNDKGBALDKGNLYTDSKLYLFGKDKET 385
DB 546 AKTIWEDYSLRNIVKELKTYKNLISKIENLEIETVLE--NIKDBEOLP-----EKIT 598
QY 386 GEDEFEVBAIKEDGSMFLIDTKPVNLSMPKNYFNPSPKSNKIYVRNPEFYLRGKISDKGF 445
DB 599 KQENKPDKILVSDLVKQVQVKVLL---MNKIDELKQTLAKVY--LKNHIVPNST 653
QY 446 NWEIAYNESVDNYLIYGDLAHIDNTRDENIKL-----NVK----- 480
DB 654 KQENKQE---PYLLIVLAKGEIDKLVFMPKVESLIINEKKNIKITEGQSDNSEPSTEGEI 709
QY 481 -----DGDIMQGMQDYK-----ANGFPDQV 501
DB 710 TQOATTKPQQAQSGALSGDSVOAQAQOQKQAOFPVPVVPBKAQVPTPPAPVNNKTENV 769
QY 502 TMDGQNVYLQTVSYDLNA-----KAVGVHYQFLYDNNVKEPVNIDPKGNTSIEYADGKSV 556
DB 770 SKLD---YLEKLYBFLNTSYICHKYLVBHSHSTNNBEKILQYKITYKESBSKLSCDPLDL 826
QY 557 FFIN-----DKRNGPDC--EIQBOHIYINGKEYTSFNDIKQIIDTKIINKIYV 603
DB 827 FNIQNNIPWYSMFSLNLSLQLENEIYEKEVNCULYKQKNDKIKNLLEBAKQVSTSV 886
QY 604 KOPARNTTVKBFILANDTESEVSELKPHRYVTVTIIONGKMSFTVSEEDFILPYKGELEK 663
DB 887 KTLSSSSMQPLSLITQDKEPVSANDTSHSTNLNLSLAKIFENILS-----LQAKKA 937
QY 664 GYQPDGWEISGFEK-----KQAGVYINLSKDTFIKVPFKIBKKEENKPTF--D 713
DB 938 IYQ---ELIQKSSSENFYEKILKQSDTYNESFTNFKSKADIDINSINDESKREKLEED 993
QY 714 VSKKQNDPVNHSQLESBHRKEDLQREBHSQKSDSTKQVTAATVLDNNISK--STNNP 771
DB 994 INKIKKTIQLSFDLTKKYLKLERLFDKCKTVGKTYMQIKCLTLLKQLEBSKLSLNNP 1052

```



```

Db      1632 AKIQKANPNRALSETTITYADSRVVEDGRSTQAVLMSALDGFNIIRYQVFTFRKNDGGA 1691
Qy      361 IDKGNLVTDSSKLVLFQKDKKEYTGSDKFNVEAIKEDGSMFLFDTPKVNLSMDKNYFNP 420
Db      1692 IDKGNLVTDSSKLVLFQKDKKEYTGSDKFNVEAIKEDGSMFLFDTPKVNLSMDKNYFNP 1751
Qy      421 SKSNKIYVRPPEFLRKISIDKGGFNMELRVNESVUNYLLYGLHIDNRPENIKLVNK 480
Db      1752 SKSNKIYVRPPEFLRKISIDKGGFNMELRVNESVUNYLLYGLHIDNRPENIKLVNK 1811
Qy      481 DGDIMWGMKDYKAMGFPDKVTMDGNVYLLQTYGSDLNAAKVGHYOFLYDNVPEVNID 540
Db      1812 DGDIMWGMKDYKAMGFPDKVTMDGNVYLLQTYGSDLNAAKVGHYOFLYDNVPEVNID 1871
Qy      541 PKGNTSIEYADGKSVVNINDKRNNGPDGEIOBHIIYINGKYSFPNDIKOITDKTLNIK 600
Db      1872 PKGNTSIEYADGKSVVNINDKRNNGPDGEIOBHIIYINGKYSFPNDIKOITDKTLNIK 1931
Qy      601 IVVDFPARNITVKEFLINKDTGSEVSELKPHRVVTYTIQNGKMSSTIVSEEDFILPVYKGB 660
Db      1932 IVVDFPARNITVKEFLINKDTGSEVSELKPHRVVTYTIQNGKMSSTIVSEEDFILPVYKGB 1991
Qy      661 LEKGYPDGEWISGFEKGKAGVYINLSKDTPIKPVFKLIEKKKEENKPTFDVSKKKDN 720
Db      1992 LEKGYPDGEWISGFEKGKAGVYINLSKDTPIKPVFKLIEKKKEENKPTFDVSKKKDN 2051
Qy      721 POWNHSLMNSHRKEDLOREHSHQKSDSTQVTAIVLDKNNISSKSTNNPNK 773
Db      2052 POWNHSLMNSHRKEDLOREHSHQKSDSTQVTAIVLDKNNISSKSTNNPNK 2104

RESULT 2
US-09-107-433-3169
; Sequence 3169, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085131
; FILING DATE: May 12, 1998
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Arinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

```

; MOLECULAR TYPE: Protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: m1ec feature
; LOCATION: (8) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169

Query Match      75.7%; Score 3048; DB 4; Length 637;
Best Local Similarity 99.3%; Pred. No. 5,66-191;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy      185 LAFAGDMRLPYKNDQKAKIKIMPEKIKETKSEFYVSSYGAVIIEGSDLSKRNKDN 244
Db      15 LAFAGDMRLPYKNDQKAKIKIMPEKIKETKSEFYVSSYGAVIIEGSDLSKRNKDN 74
Qy      245 LTKHSGKIYSDSKQOYLKDNIIILKGYALKYTYNPGKTDMLGNGVYSKEDIKIQ 304
Db      75 LTKHSGKIYSDSKQOYLKDNIIILKGYALKYTYNPGKTDMLGNGVYSKEDIKIQ 134
Qy      305 KANPNRALSETTITYADSRVVEDGRSTQAVLMSALDGFNIIRYQVFTFRKNDGGAIDKD 364
Db      135 KANPNRALSETTITYADSRVVEDGRSTQAVLMSALDGFNIIRYQVFTFRKNDGGAIDKD 194
Qy      365 GNLVTDSSKLVLFQKDKKEYTGSDKFNVEAIKEDGSMFLFDTPKVNLSMDKNYFNPBSKN 424
Db      195 GNLVTDSSKLVLFQKDKKEYTGSDKFNVEAIKEDGSMFLFDTPKVNLSMDKNYFNPBSKN 254
Qy      425 KIYVRNPEFLRKISIDKGGFNMELRVNESVUNYLLYGLHIDNRPENIKLVNKDGI 484
Db      255 KIYVRNPEFLRKISIDKGGFNMELRVNESVUNYLLYGLHIDNRPENIKLVNKDGI 314
Qy      485 MDKGMKDYKAMGFPDKVTMDGNVYLLQTYGSDLNAAKVGHYOFLYDNVPEVNIDPKGN 544
Db      315 MDKGMKDYKAMGFPDKVTMDGNVYLLQTYGSDLNAAKVGHYOFLYDNVPEVNIDPKGN 374
Qy      545 TSIEYADGKSVVNINDKRNNGPDGEIOBHIIYINGKYSFPNDIKOITDKTLNIKIVK 604
Db      375 TSIEYADGKSVVNINDKRNNGPDGEIOBHIIYINGKYSFPNDIKOITDKTLNIKIVK 434
Qy      605 DPARNTVKEFLINKDTGSEVSELKPHRVVTYTIQNGKMSSTIVSEEDFILPVYKGBLEKG 664
Db      435 DPARNTVKEFLINKDTGSEVSELKPHRVVTYTIQNGKMSSTIVSEEDFILPVYKGBLEKG 494
Qy      665 YQPDGEWISGFEKGKAGVYINLSKDTPIKPVFKLIEKKKEENKPTFDVSKKKDNPOVN 724
Db      495 YQPDGEWISGFEKGKAGVYINLSKDTPIKPVFKLIEKKKEENKPTFDVSKKKDNPOVN 554
Qy      725 HSQALNSHRKEDLOREHSHQKSDSTQVTAIVLDKNNISSKSTNNPNK 773
Db      555 HSQALNSHRKEDLOREHSHQKSDSTQVTAIVLDKNNISSKSTNNPNK 603

RESULT 3
US-09-107-433-4771
; Sequence 4771, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4771:
SEQUENCE CHARACTERISTICS:
LENGTH: 1529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1..1529
SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
US-09-107-433-4771
Query Match 18.6%; Score 750.5; DB 4; Length 1529;
Best Local Similarity 79.7%; Pred. No. 1.8e-40;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;
QY 1 KLGIASKPKFNLGNGKSGSLKKDTGVGHHQENESIKESFTIDNISTIDPFRK 60
DB 1336 KLGIIPKPKFNLKKNVNDLANKETAEVBNLLVNOISLGKSLNINHTITIDPFRK 1395
QY 61 DLKLLIKKKFVDDFTSBTGRMEYDYKYDDKNITAYDDGTULEYETKLDISKKI 120
DB 1396 DLKLLIKKKYQEDDFVN-GGRTVTERDYKYDDKNITAYDDGTULEYETKLDISKKI 1454
QY 121 YGVSPSGDGHPIIKGISNVSKNAKYVGNVYKSIETKATYDPSHKMTEDLYANIND 180
DB 1455 YGVSPSGDGHPIIKGISNVSKNAKYVGNVYKSIETKATYDPSHKMTEDLYANIND 1514
QY 181 IVDGLAF 187
DB 1515 IVDGISF 1521
RESULT 4
US-08-961-083-68
Sequence 68, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-68
Query Match 15.3%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.9e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 657 YVGELEKGYQFDGWEISGFGKDDAGYVNLNPKDFIKVFVKLEKGBENKPTDVSK 716
DB 1 YVGELEKGYQFDGWEISGFGKDDAGYVNLNPKDFIKVFVKLEKGBENKPTDVSK 60
QY 717 KDNFQVNSQLNBSRKEDLQREBSQKSDSTQVATVLDKNNISKSTNNPNK 773
DB 61 KDNFQVNSQLNBSRKEDLQREBSQKSDSTQVATVLDKNNISKSTNNPNK 117
RESULT 5
US-09-536-784-68
Sequence 68, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match 15.3%; Score 615; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.9e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

657 YKGLKGYOPDGRIGSGFGKQAGYINLSKPTFCVPRKIEKKEEENKPTPDVSK 716
1 YKGLKGYOPDGRIGSGFGKQAGYINLSKPTFCVPRKIEKKEEENKPTPDVSK 60

717 KQNPQVNHSQLNHSRKEQLQREHSQKSDSTKQVATVLDKNNISKSTNNPNK 773
61 KQNPQVNHSQLNHSRKEQLQREHSQKSDSTKQVATVLDKNNISKSTNNPNK 117

RESULT 6
US-09-200-6508-7
Sequence 7, Application US/092006508
Patent No. 6680195
GENERAL INFORMATION:
APPLICANT: Patti, Joseph M.
APPLICANT: Foster, Timothy J.
APPLICANT: Hook, Magnus A.O.
APPLICANT: Eidhim, Delidre M.
APPLICANT: Perkins, Samuel L.
TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
FILE REFERENCE: P06283US2/BAS
CURRENT APPLICATION NUMBER: US/09/200,6508
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066,815
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: 60/098,427
PRIOR FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1166
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-200-6508-7

Query Match 5.0%; Score 203; DB 4; Length 1166;
Best Local Similarity 21.6%; Pred. No. 6.8e-05;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

33 QENESIKE-KSEFTIDRNISTIRDFENKDLKIKKKF-----EVDPTTS 78
226 KNNBKKELVRNDSMTDSTKPAVAPKAVNAKMPAVAAQPAVASNNVNDLIK 285
79 EYGRMEBYDYKDDKGNIIAYDDGTDLEETE-----KDEIKSKIKYGLS 125
286 VTKGTIK--VGGCKNVAAAHDKDIEYDEFTIIDKKYAKKQDMTINIKRVNIPSDLT 341
126 PSKO-----GHPEILGKISNVSNAK---VYGNKYKSIKATYDPHAK----- 168
342 DKNDPIDITDPSGEVIAKGTEDKATKQITFTFDVVDKEDIKSLTLTYSYIDKKTVPMB 401
169 ---TMTF-----DLVANTN-DIVDELAFAQD---MRLPVK-DNDQKAEIKIRM-PEKIK 214
402 TSLMIFPRTAKKESQNTVDYQDPVH-GDSNIGSITKDEBDKQTEEQIYVNPPLKKS 460
215 EYKSEY---PYVSSYGNVIEIGBGD--LSKKKPDNLTYM-----ESGKIYDSERQO 261
461 AATNKVDIAGSQVDYGN-ITLNGSSTIIDQNTKEIKYKVNSDQQLPQSNRIYDSQYED 519
262 YLLK-DNIIILKAGYALKATTYTNPKTMDLBNNGVYSKEDIKIQANENLALSTETIYA 320

520 VTSGFDN---KKSFSNNVATIDFG-----DINSAYI-----IKVSKTTPS 558
321 DSR-NVEDGRSTQVLSALDGFNIRYQVTFPKRNDK-----E 359
559 DGBELDIAGTSMRTT--DKYGVYVAGYSNFIIVTSNDTGGGQGVTKPBEKLYKIGDYVBE 616
360 AIDKQNLVTDSKILVFGKDKKEYGDKXNVALKRDSMLPFDTPRVNLSMKNY-F 418
617 DVDKDG-----VQSTSKKCPMANVL-VLTYPDGT-----TQSVRTDAGHYEF 660
419 NPSKSNKIYV---NRPFFYLRGKIS-----DKGCFMWEJRVNESVVDNYLIYDGLHID 468
661 GGLADGETTYTKFETPTCYLPTKNGITDGERKDSGSSVYTKLNK-----DDMSLD 712
469 -----NTRDFNFKILNVKG--DINDWGRKYKANGPDDKVTDMQGNVYLQGYSDL 517
713 TGFYKPEPKYINLADYWEEDTNKQIQDANBPGIKDKVLT-----LKDSTGKV-IGTTTDA 766
518 NAKAVGVHYQPLNVNKRPNVVIDPKGNTSIEY---ADGKSVEN--INDKRNNGFDEIQ 572
767 SKR---YKF-----TDLN-NGNTYVEFETPAGTTPVKNTTADKDSNGL----- 807
573 EGHYINKEEYTSFNDIKQIIDKTLNIRKIVKDFARNT--TWKEPI--LNKDTGRVSEL 627
808 -----TTGVVIDADANMTLD-----RGFYKTPKYSIGLDVYWDNSNKGKQDSTE 851
628 KPRH-VYTTIONGR-ENSSITVSEEDPILPYKGLEKGYOPDGRIGSGFGKQAGYI 685
852 KQIKDVITVLLQNEKEVEIGITKDEN-----GR-----YRFD----- 883
686 NLSKOTFIPKPKIEKKEEENKPTPDVSKKQNPQV-----NHSQ--LNESHKREDIQ 738
884 NLDGSKT-KVIFPKPAGITQVYTTNTED-DKADAGGEVDVITTDHDFTLNGYFEEDTS 941
739 REHSQKSDSTKD 751
942 DSDSDSDSDSDSD 954

RESULT 7
US-08-973-462-8
Sequence 8, Application US/089734628
Patent No. 6191270
GENERAL INFORMATION:
APPLICANT: DROTHME, PIERRE
APPLICANT: DAUBERSTES, PIERRE
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
FILE REFERENCE: 0660-0125-0 PCT
CURRENT APPLICATION NUMBER: US/08/973,4628
CURRENT FILING DATE: 1998-02-06
EARLIER APPLICATION NUMBER: PCT/FR96/00894
EARLIER FILING DATE: 1996-06-12
EARLIER APPLICATION NUMBER: FR 95/07007
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 1786
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
US-08-973-462-8

Query Match 5.0%; Score 200.5; DB 3; Length 1786;
Best Local Similarity 20.5%; Pred. No. 0.00018;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

6 ABS--KFNK-LANGREGSLKKOT-----TGVVEHQENE-----ESIK 40
868 AESTVTPNIIIEIGQENTITTDITBEKLEBLHENVLSAALNTQSEERKGVIVDVBK 927

```

QY 41 EKSSPTI-----DRNSTIRD-PEN-----KOLKLIKKKPRENDPTS 78
DB 928 BEVATTLITVEOAEKRSANTLITRIFENNEAENVAENVAELKLNTRFNTVTLDKVE 987
QY 79 ET-----GKAME-EYDYKY-----DDKG---NIAYDSTDLLEYET-----EKLD-- 114
DB 988 ETVEISGBLENNEHDKAFPSRIFDVVKIGQENLTL---GMPRSIFTSIVIOSEKVDLN 1044
QY 115 -EIKSKIYVLSPSKDGHEILIKISNVSRNAVY-----YGNKYKSIETKATYDFH 166
DB 1045 ENNVASSILINIKMKKG---LANKLENISTSGVETVEHVRQNYVVDVDAK--- 1097
QY 167 SKTMFIDLAMININDIYDLAFAGMRLPYKNDQKKAETIKIRPEKIKETKESPYVSSV 226
DB 1098 -----DQVLGILNARGL---KEMFVLEDFVPSBSVDYVEIKDEPVKVEKET 1146
QY 227 GNVIELGSD-----LSKNRPNMLTKM-----ESGKIYSDSEKQOYLKONILKGYALK 277
DB 1147 VSIIEMENIYDVLEEKEDLTDKMLDAVEBSIEISSKESKETSIIKOK---EKQVSLV 1203
QY 278 VTTNPG-----KTMMLBGNVYSK---EDIAKIQKANPVL----- 310
DB 1204 VEBOVDNDDESVEKVLBLNNEBELMKDAVEINDYTKLIEETOBINEVADLIDOMEX 1263
QY 311 -----RALSETTIYADSRNVEDGR---STOSVLSALDGFNIIRYQVFTFPMNDKGEALDK 363
DB 1266 LKELKALSE-----DSKEIIDAKDPTLEKVIIEED-----ITTLIDVVELKDV 1309
QY 364 DGNLVTSSKLVLFQDKDEYTGEDKFNVEBAIKEDSGMLFIDTKPVNL-----SM 413
DB 1310 EEDKIEKVDL-----KOLBEDILK---VKEIKELSEILIEDYKBAIKETIDILBEKKEI 1362
QY 414 DKNYRPSKSNKIYANRPFYLRGKISDGKFMELRVNESVVD-----NYLYGDL 465
DB 1363 EKHDEKFEKEEBAEIKDLADILKEVS-----SLVEBEKQLEVEHLEKEVEHIIISGA 1417
QY 466 HINDTRDFNIKLVNDK---GDIM-----DWGKQDYKANGFPDKVTMDGNYVLTQYSDLM 518
DB 1418 HIKGJBEEDLE-EVDDLKSGIILDMKGMELGMDGESLEDYTKGGERV---BSLKDV 1473
QY 519 AAVGVHYOFLYDN-----VKPEVNIDPKGNTSIRVADKGSVFNINDK-RN 564
DB 1474 SSALGDBEOMKTRKKAQRKLEEVLLKEEVEKEEPKKIT-----KKKVFIDIKDKPKD 1528
QY 565 NGFDEIIOBHLYNKG-----EYTSNNDIKQIIDKTLN-----IKLVNDPFR-N 609
DB 1529 EIVEVEMKDEDIBBDEVEDIEEDIEEDKVEDIDEDIDEGEDKDEVIDLIYKEREIK 1588
QY 610 TTVKEFIILNKDGE-VSEELKPH-----RVTYTIQNGKMSSTIVSEED 651
DB 1589 VAKKKKLEKCYBEGVSGIKGVDEVMKVQKIDKEVDKESKALSKNDVTNVIAKQND 1648
QY 652 FLPPYKGLKGYOPDGMWISGFGSKDAGYVINI-----SKDTFIKPVK 698
DB 1649 PFSKY-KNVVKYKVFAPAFISAVAA---PASYVVGFTSLSFSSCVTIASSYTLISKVD 1705
QY 699 KLEBKKEENKX-----TPDV-----SKKDNQPVNSQUNESRKEEDLQR 739
DB 1706 TINKNKE---RPFYSVFDFIPLNKLKHYLOOMKEKESKEKNNTYEVTL-NKEKKGNVQV 1760
QY 740 EHSQKSDSTKDVATVLDKKNISKSSTNNPK 773
DB 1761 TMTKTEK-----TKVDKNNKVPKGRRTQKSK 1786

```

RESULT 8
US-09-328-352-5821

; Sequence 5821, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS

```

; FILE REFERENCE: GNC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5821
; LENGTH: 2504
; TYPE: PRN
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-5821

Query Match      4.8%; Score 191.5; DB 4; Length 2504;
Best Local Similarity 19.5%; Pred. No. 0.0011;
Matches 198; Conservative 130; Mismatches 346; Indels 341; Gaps 54;

QY 5 IAEKRNKLGNGKSGIKDQTTGVHHHQBESIKESPTID-----RNISTIDPE 58
DB 1365 LKDSKFDV---TYGVTTLATNGL-----TIKGPSTIYEGINAGSKQITNVADGI 1412
QY 59 NKOLKLIKPPRBYVDFTSENGKMBEYDYKYD-----KGNIAVDSTGLE-- 107
DB 1413 N-AKDAVNS--QDNLAKQ-NATDDAAVKTDDAKTKDKVTLKQ---DETVIDNV 1462
QY 108 -----YETEK-----LDEIKSKIYG-VLSP-----SKDGHFEILGX 137
DB 1463 KAGHISTSKBAVNSQIHKISINSIGNTVVPDGSLTNNIGTGKNNINDALISE 1522
QY 138 ISNVSKAK-VYGNNTYSIEIK-----AKTDFHSK-----TMFIDLAMINDI 181
DB 1523 VKNTATKATYTVBGN---IVKETVNRKDSNRYEATKQDVLNSVTTGDTVLANNGL 1579
QY 182 V-----DGLAFAGMRLPYKNDQKKAETIKIRPEKIKETSEYVSYGANTV 230
DB 1580 TTYDGSITKQGINAGSKQITNVADGINAKDAVNVQDLTYKERNL---GRITDNNQND 1637
QY 231 -----ELGEGDLKSNKPNMLTKMESGKIYSDSEKQOYLKONILKGYALKVTTYNPKTD 287
DB 1638 AKKDLGNQIADTKNLNDAKQDGNQITDNTVTLANTTKDQ---TYQINDTKTE 1688
QY 288 MLEGN-----VSKEDIAKIQKANPRLR-----ALSETTIYA 320
DB 1689 LNNVTGNTELENSKIDSTTELENGKLNPAAGSNDVHRKLGKRLNIIGGAASTEVAK 1748
QY 321 DS-----RNVEDGRSTOSVLSALD-----GFNIIRYQVTFPMNDKGEALDXG 365
DB 1749 TSGENYITTTQDGIQELKDSKFPSTYTGNTTATNGLTIR---EGPSVTGEGINAGG 1805
QY 366 -----MLVYDSS-----KLVLPKGD----- 380
DB 1806 KKITNVADGVNAKDAVNSQDNLMAAKQNAATDDAAVYDDAKTKDKVTLKQKQGTVLNV 1865
QY 381 -----DKB-YTGEDKRVN-----AILEDGMLFIDTKPVNLSMDKXVFN 419
DB 1866 KAGHISTSKBAVNSQIHNSINSIGNTVVPDGS---LTTNNIG-GTGKNNIND 1921
QY 420 -----PSKSKIRYVRNPEFYLRGKISDGKFMELRVNESVVDVYLLYGDHLINDR 471
DB 1922 AISEVGTATKATYTTGEGNIVYKGTANQGSTNVEVSTKQDVLNSVTTGDTVLANN 1980
QY 472 DFNILKLVNDGDIMGMDKYKANGFPDKVTMDGNAVLTQYSDLMNAKAVGVHYOFLYD 531
DB 1981 -----GLTIKQPSI---TKD-GVNAAGKKTIDVANKVIAQNSKDAV---GAOVHHISN 2028
QY 532 NYKPEVNIDPKGNTSIRVADKGSVFNINDKRNNGFPGIIOBHIIYNGKEYSFNDIKQ 591
DB 2029 STNKSIG---GNTVNA-PDGSILTNNGIGTGKNNINDAKS---VDEKVTNGVNDLTQ 2079
QY 592 IIDKTLNLIKLVNDPARN---TYKEFIILNKDGEVSEELKPHRYVTYTIQNG-----KX 641
DB 2080 -----KGLN-----FANDQKTTQGAVVRKGLD-----TINIVGADAKTAEK 2120
QY 642 MSSTIYSE-EDPILVYKGLKGYOPDGMELISG-----EGK 678
DB 2121 SGENIITRTTEBDV---KLEMLKDVKFSVNVGAVLNDQGLIKGSPSTYNGINAGG 2176

```

Qy	Db	Qy	Db
679	2177	724	2233
UDAGV---INLSXOFIK-PVFKKIEBKEEENKFFDVSK-----KKDNQV 723	KQINNVADGIN-ADDAVNGQGLQKQINEVMDQIGK---DIEKSDHVAQYDKDKKGVDK 2232	NHSQLNBSHREKDIQREBHSQKSDSTQVTV-----ATVLDKNNTSSKSTNN 770	NSVTLGGGEKKTINLVADGCVAAEGSKDAVNGQGLMNIQNVQDKNSDNIKNIQNN 2287

RESULT 9
HS-09-107-A33-A976

; Sequence 4976, Application US/09107433

Patent No. 6800744
GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID

SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC PURPOSES

INTERFOLIOS
NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS

STREET: 100 Beaver Street
CITY: Waltham

STATE: Massachusetts
COUNTRY: USA

COUNTRY: USA
ZIP: 02354

```

;      COMPUTER READABLE FORM:
;
;      MEDIUM TYPE: CD-ROM ISO9660

```

COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>

```

; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA.

```

APPLICATION NUMBER: US/09/107,433

FILING DATE: 30-JUN-1998
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/
FILING DATE: May 12, 19

APPLICATION NUMBER
FILING DATE: July

ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Paolo Donato

NAME: ALFONSO, Pamela Denise
REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 893-5007
TELEFAX: (781) 893-8277

```

1      INFORMATION FOR SEQ ID NO: 4976:
2      SEQUENCE CHARACTERISTICS:
3          LENGTH: 670 amino acids
4          TYPE: amino acid
5          TOPOLOGY: linear
6          MOLECULE TYPE: protein
7          HYPOTHEetical: YES
8          ORIGINAL SOURCE:
9              ORGANISM: Streptococcus pneumoniae
10         FEATURE:
11             NAME/KEY: misc feature
12             LOCATION: (B) LOCATION 1...670
13         SEQUENCE DESCRIPTION: SEQ ID NO: 4976:
14         IS-09-107-433-4976

```

```

Oy 316 TTIYADSNVBDGASTGVLSALADGPNIRIQRVTFERNKNDGEBALDIXGNLTJDSKLY 375
Db 165 ATI-----LGNINKNPSIVFT-----GLTDDCAQV 193
Oy 376 LFG-KDKEYTEGDFNFBALKEDGSMFLPDTKPVNLN--DKVFNPSKNPIYVNPB 432
Db 194 EMGFTEDISYSGITIDMGNALNBERG-----TKAKPLPLINSSGAPALGNNVVTIKNVT 247
Oy 433 F---YLREKISDKGFPNMLERPNBESVNDNYLYKGLHIDNTRDPNIKLVKQSDIM---D 486
Db 248 FKDSYQGHAIQIAGSKN-----VLVDSRFLQA-----LPKTKRQDIISKES 291
Oy 487 WGMKDYKANGPPDKYTDMDG---NVYLTQGY---SDLNAK---AVGVHYQFLYDVKPE 536
Db 292 IQIBFLTRKGFPYALND-DGKKSENVTIQNSYFGSKDPSKGEVLVATIGHYQTLSQNP SN 350
Oy 537 VNIDPKGNTSLBYADGKSVFNPNINKRNGPBGKIQBGHIY-----INGEYTSND 588
Db 351 IKILNNHPDNMMYTAGVRFTGFTDVLIKGNRFPFKKGVSEVHRESGALVMAVSNTTKD 410
Oy 589 I---KQII-----DKTLNKIYVKDPAK-----NTYKESFILNKDGYEYSLK 628
Db 411 LLDLNTKQYVIAENIFNIDPPTKALR-VAKDSAEYLGKVSIDITVTKNVLNNKERTEQ-- 467
Oy 629 PHRYVTVTIQNGKMSSTIV--SEEDFILPVYKGEI---EKGYQDFGMEI---SGPECK 678
Db 468 FNIBFLRYSDWLVVSENSIPGGKBEIVLEDSKGIIVLNNQFYNLSGKYSISFKSNANAK 527
Oy 679 K-----DAGVILNSKDTFLKPVFKKIEKKKEE-----NKPTFDVSK 716
Db 528 EPVIRDSISGNFNI-VTENGILKYLIVNNLSDKNKKKKNKEBKQYNSNNVYIDSNQKNBPN 586
Oy 717 KQDNPNQVNSQNLNESHK-----EDLQREHSHQSKSDTKOV 752
Db 587 SKDNRNQNDKIDNKQDNKTEEBVNYKIVGDRGTEHNNHNSKSIADV 632

```

RESULT 10
MS-09-583-110-3019

Sequence 3019, Application US/09583110

; Patent NO. 6699/03
; GENERAL INFORMATION

APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relat

TITLE OF INVENTION: Pneumoniae for Diagn
PAT REFERENCE: PAT#00-072

FILE NUMBER: FHH00-01A
CURRENT APPLICATION NUMBER: US/09/583,111
CURRENT FILING DATE: 0000-00-00

; CURRENT FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/107,433

; PRIOR FILING DATE: 1998-06-30
 ; PRIOR APPLICATION NUMBER: US 60/085,131

PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: IIS 60/051,553

PRIOR FILING DATE

; NUMBER OF 3
; SEQ ID NO 30

```

; LENGTH: 708
; TYPE: PRT
;

```

```

: ORGANISM: Streptococcus pneumoniae
US-09-583-110-3013

Query Match      4.7%; Score 188.5; DB 4; Length 708;
Best Local Similarity 21.4%; Pred. No. 0.00031;
Matches 138; Conservative 92; Mismatches 233; Indels 183; Gaps 33

```

QY 223 VSSYGAVNIKLGGBDLKKPKPPNLTVMESGRTYSDEKQQY---LKNDNTILR--KGVALK 277
 : :
 Db 92 ISSNGTI-----RSNSQLDNRT-VBS-TYTSTBNKNSYEDVLSDRLLIKKEFEPTALS 142
 : :
 QY 278 VTYYN-----PCKTDLRGNGVYSKEDIATIQ---KANPNLRALSE 315
 : :
 143 VKDVGAGWDGIHDDRQAIGDAIIDAAOGLGGGNVYFPFGTYLVKELVFLKSTHTLELNK 202

```

QY 316 TTIVADSRVBDGRSTQSVLMSALDGFNIIRYQVFTPKMDKGBALDKOGLVTDSSKLY 375
DB 203 ATII-----LNGINIKNHPISYFMT-----GLTTDGAQV 231
QY 376 LFG-KODKEYTGBDKENVEAIKEDGMLPDTKPVNL-SM--DKNYNPSKSNXYVNPZ 432
DB 232 BWGPTEDISYSGGTIDMKNALNEEG-----TKAKULPLINSGCAFALGNSNNVTIKNVT 285
QY 433 F---YLKGLSDKGGPNMELRVNESVVDNLTLYGDHIDNTRPNIKLVNKGDDIM---D 486
DB 286 FDSYOGHAIQIAGSKN-----VYVDSRPLGQA-----LPTKDKDQIISKES 339
QY 487 WKKKDYKANGFPDKVTDMDG-----NVYLTQGY-----SDLNK-----AVGVHYQPLVDNKPZ 536
DB 330 IQIEBPTKRGFPALND--DQKSENVITIQNSYFGSKDSELYALIGHYQTLSTQNPEN 388
QY 537 VNIDPKGNTSIBYADKSVFVFNIDKRNNGPDGEIOGHY-----INGKEYTSFND 588
DB 389 IKLINHPNNMYAGVRFTGFTDVLKGNRFDKKVGESVHYRESGALVNAVSYKNTKD 448
QY 589 I-----KQII-----DKTINIKIIVNDPAP-----NTYKKEPIIAKDGSELSK 628
DB 449 LLDLNQVAVIENIFNIADPKTKAIR-VAKDSABYLKQVSDITVTKVNNNSKETBQ-- 505
QY 629 PRRVTVTIOGKEMSSYIV--SESDPLPYKGL-----EKGYQPDGMBI-----SGFBGK 678
DB 506 PNIHLRVSDNLVVSERSITFGKGIYIBSKKITVLANQVFNLSGKTSPIKSNANGK 565
QY 679 K-----DAGVYINLSKDTPIKPFKIKBEKKEE-----NKPTPDVSK 716
DB 566 BPVIRSDGPNFI-VYENGLYKIVTNLSPDKNEKKEKKEKQYNSNNVIDISNKGKGFNS 624
QY 717 KXDNQPVNHSQLSBHRK-----EDLQREHSQKSDSTKV 752
DB 625 SKDNQKNDKIDNKQDKTEEVNYKIVGDGRETENHINKSEIADV 670

RESULT 11
US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matlare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Berkreiser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-728-470-10

Query Match 4.7%; Score 187.5; DB 2; Length 1529;
Best Local Similarity 20.1%; Pred. No. 0.001;
Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;

QY 9 KFKUL---GNGKESLKKDTTGVGHHHQBENBSIKESSTFIDRNISTRIDFENKDLK-- 63
DB 531 RKNVNSINGTK-----GLKFLANQNNFTHKPDGELNIS-GIVTINGTKKQVYM 580
QY 64 KLIKKEFREYVDPTSETGKMEBYDY-KYDDKANIAYDGDGTELEYTEKLDIYSKIYV 122
DB 581 NAKSDSYMNVSSLTNT---VQKFTFIKFDVSGS-----NSODLSSRRSPAG 625
QY 123 VLSPSKDGHE-ILKGIS-NVSGIAKYV-----GNNTKSIIE--KATKDPHSKT 169
DB 626 V-----HNGIGGKTNFNIGANAALFKLPNAATDPKKEPLITTANANITATGSDSS 678
QY 170 MTEPLYANIND-----IVDGLAFAGDWRLFVKNDQKKAIRIKIMPEKIKETSEYPVS 224
DB 679 VMFDIHANLTSRAAGIMDSINITGIDPSITSNRNSNAFEIKKDLINATGNSPS-- 735
QY 225 SYGVNIELEGSDLSKPKDNLTYQESGKIYSDEKQOYLKDNITLAKGALKYTTNPG 284
DB 736 -----LKQTKDSFYNEYSKHAINSHNLT----- 760
QY 285 KTDMLBGNVYSKEDIKAKIPANPLALSETTIYADSRVBDGRSTQSVLMSALDGFNI 344
DB 761 ---LGSNVTLLGGENSSSITGANININYNKANVTLQADTNSNTGLKAKRTLLG----- 809
QY 345 IRYOVFTFRKNDKGBALDKOGLVTDSSKLYLFGK---DKKEYTGBDKENVEAIKEDGS 400
DB 810 -----NISVGNLSLTGANANINIGNLSIADSTFKGE----- 841
QY 401 MLEPDTKPVNLSDKNYFNPSKSNKIYVRNPEFLKGISDKGFGNMLRVNESVVDNLT 460
DB 842 -----ASDMLNITGFTNNGTAN-INIKQGVVLAQGDINNKGGLN--ITTNASGTOKTI 892
QY 461 IYGDHIDNTRDNIKNLVK-DGDIMDWG-----MKDYKAN-----G 496
DB 893 INGNI-TNEKGDINIK-NIRADABITQIGNISQKRGMLTSSDVTNITNOITTAQYEGG 950
QY 497 FPDKVTMDGNVYLTQ-----GY-----SDL-----NAKAVGVHY 526
DB 951 RSDSSBAKANLVIQTEGLADGMINSGFNKAETAKNSDDLIGNASGNADAKV-- 1008
QY 527 QPLVDNYPEVNITDPKNTSIBYADKSVFVFNIDKRNNGPD--GBIOBOHIIYNGKEYT 584
DB 1009 --TFDKVY-----DSKIST-----DGHNVITNSEVKTSNSNGNNSGTGLTISADV 1056
QY 585 SPNDIQIIDKTINIKIIVNDPAPNTVKE-PIIANOTGSEVSKPFRVTVTIOG---- 639
DB 1057 VNNNVTS--KKTINISAA---AGNVTTKEGTTINATTSVY-----VTAONGITIG 1102
QY 640 --KEMSSIVSEDPFI 653
DB 1103 NITSQNVTVATANTLV 1118

```

RESULT 12

```

US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreaser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-10

Query Match          4.7%; Score 187.5; DB 3; Length 1529;
Best Local Similarity 20.1%; Pred. No. 0.001;
Matches 148; Conservative 94; Mismatches 255; Indels 239; Gaps 37;

QY 9 KKKRL---GNGKESGSIKKDTGVHGHQESIKKESFTIDRISITIRDPENDLK-- 63
DB 531 KFNVNVSINGTK-----GLKFIANQNNPFRKFDGLNIS-GIVINQTKKDVYRW 580
QY 64 KLIKKRREVDFTSETKRMEYDY-KYDDKGNIIAYDDGDTLEYETKLEIKSIKYG 122
DB 581 NASDSYVNVSSLTNT---VQKFFIKFVDSGS-----NSQDLRSSRRSPAG 625
QY 123 VLSRSKQGHF-IIGKIS-ANSKAKVY-----GNNYSIER---KATKIDFHSKT 169
DB 626 V-----HFGIGGKTNFNIGANAKALFKLKPNAATDPKGLPTTFNANTATGNSDS 678
QY 170 MTFPLVANIIND-----IYDGLAFAGDMRLFYVNDOKKAEIKIRPEKIKETKSEPYVS 224
DB 679 VMFPIHANTSRAGINNDISINITGDLPSITSHRNANAFBIKQDLTINATGSNPS--- 735
QY 225 SYGVNILEGSDLSRNNKPDNITRMEGKIYSDSEKQVILKDNITILRGVALKVTYTPG 284
DB 736 -----LKQTKOSFYNYSKHAINSNHLTI----- 760

```

```

QY 285 KTDMLEGKGVSKEDIAKIQKANPILRALSETTIYADSRNVEDGRSTQSVLMSLDGFINI 344
DB 761 ----LGNVTLGEGNSSSSITNINITYKANVTLDQTSNSTGLKKRTLITLG----- 809
QY 345 IRYQVTFKMDKGAIDKQGNLVYDSSKLVLRG---DKREYGEDKFNVEAIKEDGS 400
DB 810 -----NIEGNSLSTGANANIVGNISIAEDSTPKGE----- 841
QY 401 MFLITKPVNLMSDKNYNPSKSNKIYVRNPEFYLRGKISDGGFVWELRVNVSVDNYL 460
DB 842 -----ASDNLNITGTFNNGTAN-INIKQGVVKLQGDINNKGGLN--ITNASTGTOKTI 892
QY 461 IYGDILINPTDFFIKLVNK-DGDIIMDG-----MKDYKAN-----G 496
DB 893 INGNITWEKEDLNK-NIKDABEIQIGNISQKGNLTISSDKNITNQITIKGVEGG 950
QY 497 PDKVTMDGNYVLTOT-----GY-----SPL-----NAKAVGVHY 526
DB 951 RSDSSEARNANLTIQTKELKLAGDLNIGFNRABITAKRGSDLTIGNAGGNADAKV-- 1008
QY 527 QELYNVNPEVNIIDKGNITSIEYADKSVFVNIIDKRNNGPD--GEIQGHYINGEYTI 584
DB 1009 --TFDKVR-----DSKIST-----DEHNVTLNSEVKTSGSSNAGDNDSTGLTISKQVT 1056
QY 585 SFNDIKQIITDKTLNIRIYVKDFARNTYVK-FILNKDTGEYSLEKPRHRTVTIQNG-- 639
DB 1057 VNNVNTS--HKTINISAA--AGNVTKBGTIVNATGSVE-----VTAQNGTIMG 1102
QY 640 --KEMSSITVSHEDPI 653
DB 1103 NITSQVTVTATNMLV 1118

```

```

RESULT 13
US-09-200-650E-3
; Sequence 3, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Richlin, Deirdre M.
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO: 3
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-200-650E-3

Query Match          4.6%; Score 186.5; DB 4; Length 930;
Best Local Similarity 20.8%; Pred. No. 0.0006;
Matches 175; Conservative 121; Mismatches 346; Indels 199; Gaps 42;

QY 3 GEIASEKPKNLGNGKESGSIK-DTGVHGHQ---ENRESIKKESFTIDRISITIRDFE 58
DB 56 GELNLSKNTBTAPSNKTKTKVDSKQLKDNQYTAADPKYTWSDSAIVKERTSSNQSFQ 115
QY 59 NKDLKLIKKRREVDFTSETKRMEYDYKYDDKGNIIAYDDGDTLEYETKLEIKSIKYG 117
DB 116 NATNQSTTKTSNVTNNDKSSITYSNE-----TDKSNLTQAKDVSTPKTTIKPRTLN 169
QY 118 SKIVGVLSPSKQGHFEILIGKISNVSKAKVYVYNNYSIKIKATKYDFHSKTMTPDLVAN 177

```

```

Db 170 RMAVNTVAAPQGG-----TNV--NDRKVFHSN---IDIAIDKGVHQTGTEBFMAT 215
Qy 178 INDI-----VDGLAPAGDMLRPVKNDQKAEIKIMPEKIKETKSEYFVSSSYGNV 229
Db 216 SSDVLAALKRANNTIDDSVKEGD--TFTEKYGQYFRPGSVLP--SGTOMLY--NAQGN 267
Qy 230 IELGEGDLSKNRP-----DNLTGMSGKIYSDBEKQOYLKDNIIIRKGYALVATY 281
Db 268 IAKGIYDSTTTTNTTTFNTVVDQYTNVRG-----SFEQYAPAK-----RKNATTDKTY 316
Qy 282 NRGKTDMLBGCNVYSKEDIQAKQKAPNLRALSETTYAD---SRNV-----EDGRST 331
Db 317 ---KKEVTLGANDTYSEELI--VDYGNKKAQPLISSTNYINNEBLSHMTAYVNOQPKTYT 371
Qy 332 OSVLMSALDGF-----NIIKYOVFTPGQNDKGBALIDKQNLVTDSSGLVLFQGDQKRYTGB 387
Db 372 KQTFPTNLTYGTFKPNPAKPKFI--YEVDQNOQFVD--SFTPDYTKL-----KQVT-- 417
Qy 388 DKFNVEALKEGDSMLPFDTRPVNLSDKNYF-----NPSKSNKLYVNPPEFYLRGKI 439
Db 418 DQFDV-IYSNDKKTATVVDLMKGQTSNKGYYIIQGVAVPNNSSTDN-----GKI 464
Qy 440 ----SDKGGFWB--LAVNESVVDNYLIYGDHLIDTRDENIKLANVQD--DIMDKGM 489
Db 465 DYTLDTDKTKYMSNSYSNVSSTAN---GPQKKYNLGDYWEEDTNKQDANEKGI 520
Qy 490 KQY-----KANGPP-DKVTMDQGNVYLOTGYSDLNAKAVGVH-----YQFLYDNK 534
Db 521 KGVYVYILKDSNKGKEDRTTIDENGKQYTGSLN-----GTVSVERSTAGCTPTTANVG 574
Qy 535 PEVNIDPKMNTS--IEVADG-----KSVFENINDK--RNNPGDGEIOPQHIIYNGK 581
Db 575 TDDAVDSDELFTTGAVIKQADNMTLDSGFYKTPKYSGLDYVWYDSNNDGKRDSTKGIKGV 634
Qy 582 ERTSNDIKQIIDKTLNITIVKQDPAKNTTVEKFLANKOTGEVSEL--KPKRYTYTQIN- 638
Db 635 KVTLLQNEKEVEIGTT-----ETDNGKTRPDNLDSGKXKVPFKPKAGLTGTGTAT 684
Qy 639 -----GKMSSTVSEEDFLPYKGBLEKGYOPDGMELISGFGKQDAGVIMLSKD 690
Db 665 TEDDDADAGGVYVITTHDDF-----TLONGYEEBTSDBSDSDSDSDSDSDSDSD 737
Qy 691 TPIKVPFKKIEBKKEENFTFDVSKKQNPVNHSQLNESHKEDLQREBSQKSDSTK 750
Db 738 S-----DSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDS 788
Qy 751 D 751
Db 789 D 789

```

```

; TYPE: PRF
; ORGANISM: Staphylococcus aureus
US-09-200-650E-5
Query Match 4.6%; Score 186; DB 4; Length 1315;
Best Local Similarity 20.8%; Pred. No. 0.001;
Matches 188; Conservative 115; Mismatches 328; Indels 274; Gaps 51;

Qy 23 KQTYGVHEHHQRE-----ESIEKESFTIDRISITIDPENQDL---KK 64
Db 164 KSVNVQPTNEERKKVADATBESTTTLNVKSDAISNDELVDNN--SUSNNENADIIILPKS 222
Qy 65 LKKKE-----REVD--FTSEYKMEBEYDYKDDGN--IIAYDDGTDL 106
Db 223 TAPKRLNTMRILAAVQPSSTBAKNVADITSNTTLVVD-----ADKNNKIVPAQDYLSL 277
Qy 107 ERIETEKLDRIKS-----KLYGLSPSKQGHFELIGKLSNVSKAKATYYGNNK 154
Db 278 KSOITVDKVKSGDYFTIKYSDTVQVYG--LMPK-----DIKNI--GDIKDPNNGB 324
Qy 155 SLEIKATKYDPSKWTTP--DLXANINDVGLAFA-----GDMELFYK--DN 198
Db 325 T--IATKEDTANNLITFTDYDRFNSVQMGINISYMDADLTIVSKRDVFNNTIGN 382
Qy 199 DQKKAERIKRMEKIKETKSEY-----PYSSYGN-----VILGEGDLSNRK- 241
Db 383 TTKKTANI QYPDYVNEKNSIGSAFTETVSHVGNKRNPGYKQTIYVNPSENSLTVAKL 442
Qy 242 -----PDNLTKMESGKIYSD--SEKQOYLKDNIIILRKG-----ALKV 278
Db 443 KVOAVHSSYVNNI--GOINQDVIDIKIYQPKYTLNKGVDNTEKELTDVTOYLOK 497
Qy 279 TTYNPKETMLBGNVYSKEDIQAKQKAPNLRALSETTYADSRNVEDRSTOSVLSA 338
Db 498 ITTGDNNSAVID---FGAADSAYVMVNTKQ-----YMS--ESFTLVQMNTLSS 543
Qy 339 LDGFNIIRYQVTFKKNNDGBA--IDKQGNIV--TDSK-----LVLPK 379
Db 544 TGNKSVSTGNALGFTNNGGAGQGEVYKIGNYWEEDTNKQVQBLERKGVNVTYVFDN 603
Qy 380 DDEKTYGEKFNVEALKEGDSMLFI-----DTEPVNLSDKNY--FNPSSKNKLYVNPPEF 433
Db 604 NNTTKYGE--AVTKEDGSTILPMLPNGDYRVERFNSLPKGEYVTPSKG--ANBEL 654
Qy 434 YLRGKISDKGFWMBELRVNESVVDNYLIYGDHLIDTRDENIKLANVQDQIMDKGMQDYK 493
Db 655 -----DSNGLSVITVNGK--DN--LSADLGI-----YKPKNML--GDYV--W--EDTN 693
Qy 494 ANGFPDK-----VTMDGNVYLOTGYSDLNAKAVGVHQL--YDNVKEPVNI--D 540
Db 694 KNGIQODEKGISGVVTLTKDENGNV--LKVVTYTDAGK-----YKFTDLONGYKVEFTT 747
Qy 541 PKGNISIEYADGKSVFENINDKRN-----NGFGEIQBQHIYNGKETSFNDIKQIIDKT 596
Db 748 PEGYTFPTTVTSGSDIEKDSNGLTTTGVINAGADMTLDSGFTKPK---YNLGNVYWBOT 803
Qy 597 LNIKIVKQDPAKNTTVEKFLANKOTGEVSELKP--HRVTTYIONGKEMSSITVSEEDFILP 655
Db 804 -----KNDGQDSTBKIGSVTTLKN-----ENGVLQ 832
Qy 656 VYKGBLEKGYOPDGMELISGF--EGKQDAGVINLSADOTIKVPFKKIEBKKEENFTFD 713
Db 833 TYKTDGKGYOFGLENGYKVEFRPSPGTYTP--QVSGGTDBRIDNGSTVG 884
Qy 714 VSKKQNPVNHSQLNESHKEDLQREBSQKSDSTKD-----VTATVLDKNNISSKST 767
Db 885 VIKQDNMTIDSGFYKPTYNLGDYWEEDTNKQVQDKEKGISGVVTLKDBNDKYLKV 944
Qy 768 TNNPN 772
Db 945 TTDEN 949

```

```

; RESULT 14
; US-09-200-650E-5
; Sequence 5, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Bidhinn, Deirdre M.
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P062830S2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650B
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1315

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:19:57 ; Search time 141 Seconds
(without alignments)
1791.327 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KGEIABSRKYNLGNKRGES.....ATYLDKNNISKSTNNPNK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues
Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCFUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	US-10-067-385-8	Sequence 8, Appl1
2	4026	100.0	2119	US-09-769-744A-28	Sequence 28, Appl1
3	4026	100.0	2140	US-10-282-122A-73670	Sequence 73670, A
4	4026	100.0	2140	US-10-472-928-1180	Sequence 1180, Ap
5	615	15.3	117	US-09-765-472-68	Sequence 68, Appl1
6	227	5.6	1196	US-10-282-122A-52737	Sequence 52737, A
7	222	5.5	861	US-09-820-843A-109	Sequence 109, Appl
8	218	5.4	1639	US-10-087-464-10	Sequence 10, Appl1
9	212	5.3	1141	US-10-282-122A-70251	Sequence 70251, A
10	207.5	5.2	4668	US-10-282-122A-76865	Sequence 76865, A
11	207	5.1	1191	US-10-282-122A-52048	Sequence 52048, A
12	203	5.0	1166	US-10-744-616-7	Sequence 7, Appl1
13	203	5.0	1881	US-10-032-585-7646	Sequence 7646, Ap

14	200.5	5.0	1786	9	US-09-742-096-3	Sequence 3, Appl1
15	200.5	5.0	1787	15	US-10-415-253-2	Sequence 2, Appl1
16	196.5	4.9	1184	16	US-10-282-122A-53254	Sequence 53254, A
17	196.5	4.9	5176	16	US-10-437-963-150986	Sequence 150986,
18	195	4.8	1849	15	US-10-637-544-2	Sequence 2, Appl1
19	194	4.8	1178	15	US-10-282-122A-52434	Sequence 52434, A
20	194	4.8	1957	15	US-10-369-493-2070	Sequence 2070, Ap
21	193	4.8	903	15	US-10-282-122A-52328	Sequence 52328, A
22	193	4.8	909	15	US-10-282-122A-52109	Sequence 52109, A
23	192.5	4.8	708	17	US-10-472-928-3792	Sequence 3792, Ap
24	191.5	4.8	1156	15	US-10-369-493-43	Sequence 43, Appl1
25	190	4.7	1789	15	US-10-282-122A-43832	Sequence 43832, A
26	190	4.7	1088	15	US-10-282-122A-46664	Sequence 46664, A
27	189.5	4.7	841	9	US-09-815-242-5779	Sequence 5779, Ap
28	189.5	4.7	841	9	US-09-815-242-12751	Sequence 12751, A
29	189.5	4.7	1143	15	US-10-369-493-11081	Sequence 11081, A
30	189	4.7	1313	15	US-10-282-122A-76863	Sequence 76863, A
31	188	4.7	943	17	US-10-741-849-7309	Sequence 7309, Ap
32	187	4.6	1103	15	US-10-282-122A-76866	Sequence 76866, A
33	187	4.6	2265	15	US-10-282-122A-45123	Sequence 45123, A
34	186.5	4.6	930	17	US-10-744-616-3	Sequence 3, Appl1
35	186.5	4.6	1790	15	US-10-369-493-1586	Sequence 1586, Ap
36	186.5	4.6	1903	16	US-10-766-993-3	Sequence 3, Appl1
37	186	4.6	1315	17	US-10-744-616-5	Sequence 5, Appl1
38	185	4.6	1385	15	US-10-282-122A-44324	Sequence 44324, A
39	185	4.6	3225	16	US-10-408-765A-254	Sequence 254, App
40	185	4.6	5005	15	US-10-282-122A-76871	Sequence 76871, A
41	184.5	4.6	1875	15	US-10-282-122A-62947	Sequence 62947, A
42	184	4.6	1875	15	US-10-369-493-22285	Sequence 22285, A
43	183	4.5	2402	16	US-10-661-809-20	Sequence 20, Appl1
44	182.5	4.5	882	15	US-10-282-122A-53247	Sequence 53247, A
45	182.5	4.5	923	16	US-10-771-931-31	Sequence 31, Appl1

ALIGNMENTS

RESULT 1
US-10-067-385-8
Sequence 8, Application US/10067385
Publication No. US20020110562A1
GENERAL INFORMATION:
APPLICANT: Chou, Gil
TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines
FILE REFERENCE: 469201-589
CURRENT FILING DATE: US/10/067, 385
PRIOR APPLICATION NUMBER: 2002-02-05
PRIOR FILING DATE: 2000-06-09
PRIOR APPLICATION NUMBER: US/09/590, 991
PRIOR FILING DATE: 1999-06-10
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8
LENGTH: 773
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-067-385-8
Query Match 100.0%; Score 4026; DB 13; Length 773;
Best Local Similarity 100.0%; Pred. No. 1.6e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KGEIABSRKYNLGNKRGESIKKQTTVEHHHONKESIKKSSFTIDRNISTIDPNK 60
DB 1 KGEIABSRKYNLGNKRGESIKKQTTVEHHHONKESIKKSSFTIDRNISTIDPNK 60
QY 61 DLKLLIKKKRREYDDPTSEGRKMEYDYKDDGKNTIAYDDGTDLEETKLEIKSKI 120
DB 61 DLKLLIKKKRREYDDPTSEGRKMEYDYKDDGKNTIAYDDGTDLEETKLEIKSKI 120
QY 121 YGVLSPSKDGHPFELIGKISNVSKNAKYVYGNMYKSIKATKYPDPHSKTMTPDLXANIND 180


```

FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282.1222A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 73670
LENGTH: 2140
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-282-1222A-73670

```

Query Match 100.0%; Score 4026; DB 15; Length 2140;

Best Local Similarity 100.0%; Pred. No. 5.9e-216; Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 KLGRIASKFRNKGKSGSLKDTTGVHHHQNBSIKKSSFTIDRNISTIRDFENK 60
DB 1334 KLGRIASKFRNKGKSGSLKDTTGVHHHQNBSIKKSSFTIDRNISTIRDFENK 1393
QY 61 DLKKLKKKFRVDDFTSEGRMEYDYKYDDKGNIIAYDDGTDLEYTEKLDIISKI 120
DB 1394 DLKKLKKKFRVDDFTSEGRMEYDYKYDDKGNIIAYDDGTDLEYTEKLDIISKI 1453
QY 121 YGVLSPSKDGFEILGKISNVSKNAKYGGNNYSIRIKATKYPHSHKTMFPDLVYANIND 180
DB 1454 YGVLSPSKDGFEILGKISNVSKNAKYGGNNYSIRIKATKYPHSHKTMFPDLVYANIND 1513
QY 181 IYDGLAFAGDMRLFVKONDQKAEIKIRMPERIKETSEYEVYSSYGAVIELGEGDLSKN 240
DB 1514 IYDGLAFAGDMRLFVKONDQKAEIKIRMPERIKETSEYEVYSSYGAVIELGEGDLSKN 1573
QY 241 KPDNLITMNSGKIVSDSEKQOYLLKDNIIILKGYALKVTYNNPGKTMLEBNGVYSKEDI 300
DB 1574 KPDNLITMNSGKIVSDSEKQOYLLKDNIIILKGYALKVTYNNPGKTMLEBNGVYSKEDI 1633
QY 301 AKIOKAMPNIALSETTIVYASRVNEDGRSTQVLSALDGFNIIIRYOVFTFRKNDKGEA 360
DB 1634 AKIOKAMPNIALSETTIVYASRVNEDGRSTQVLSALDGFNIIIRYOVFTFRKNDKGEA 1693
QY 361 IDKGNLVTSSKLVLFKQDKKEYTGSDKFNVEAIKEDGSMFLPDTKRVNLSMDKNYFNP 420
DB 1694 IDKGNLVTSSKLVLFKQDKKEYTGSDKFNVEAIKEDGSMFLPDTKRVNLSMDKNYFNP 1753
QY 421 SKSNKIVYVRNPEFLRKIGISGKSPNWEIARVNSVVDNYLLYGLHIDNTRDFNIIKLVNK 480
DB 1754 SKSNKIVYVRNPEFLRKIGISGKSPNWEIARVNSVVDNYLLYGLHIDNTRDFNIIKLVNK 1813
QY 481 DGDIMDMGMDYKANGFPDKVTDDMDGNVYLLQTYGSDLNAAKAVGVHVOFLYDNPVPEVNI 540
DB 1814 DGDIMDMGMDYKANGFPDKVTDDMDGNVYLLQTYGSDLNAAKAVGVHVOFLYDNPVPEVNI 1873
QY 541 PKGNTSLEIYADGKSVFNINDKRNNGFDGEIQBOHIYIINGKEYTSFNDIKOIIDKTLNIK 600

```

```

DB 1874 PKGNTSLEIYADGKSVFNINDKRNNGFDGEIQBOHIYIINGKEYTSFNDIKOIIDKTLNIK 1933
QY 601 IYVQDPARNTTVEKFIINIKDTGVESELKPHRYVTTITONGKMSSTIYSEEDPILPYNGE 660
DB 1934 IYVQDPARNTTVEKFIINIKDTGVESELKPHRYVTTITONGKMSSTIYSEEDPILPYNGE 1993
QY 661 LEKGQFDGWEISGFEKQDAGVYINLSKQFFIYVPFKIIEKKBBENKPTFDVSKKON 720
DB 1994 LEKGQFDGWEISGFEKQDAGVYINLSKQFFIYVPFKIIEKKBBENKPTFDVSKKON 2053
QY 721 POWNHSQLNESHKREDLQREHSSQKSDSTQVATVLDKNNISSKSTNNPNK 773
DB 2054 POWNHSQLNESHKREDLQREHSSQKSDSTQVATVLDKNNISSKSTNNPNK 2106

```

RESULT 4

US-10-472-928-1180

Sequence 1180, Application US/10472928

Publication No. US20050020813A1

GENERAL INFORMATION:

APPLICANT: CHIRON SPA

TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS

CURRENT APPLICATION NUMBER: P026926MO

FILE REFERENCE: P026926MO

PRIOR FILING DATE: 2003-09-26

PRIOR APPLICATION NUMBER: GB-0107658.7

PRIOR FILING DATE: 2001-03-27

NUMBER OF SEQ ID NOS: 4979

SOFTWARE: SeqMan99, version 1.03

SEQ ID NO 1180

LENGTH: 2140

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

FEATURE:

OTHER INFORMATION: serine protease, subtilase family

OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LpxTG)

OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.8+01)

US-10-472-928-1180

Query Match 100.0%; Score 4026; DB 17; Length 2140;

Best Local Similarity 100.0%; Pred. No. 5.9e-216; Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 KLGRIASKFRNKGKSGSLKDTTGVHHHQNBSIKKSSFTIDRNISTIRDFENK 60
DB 1334 KLGRIASKFRNKGKSGSLKDTTGVHHHQNBSIKKSSFTIDRNISTIRDFENK 1393
QY 61 DLKKLKKKFRVDDFTSEGRMEYDYKYDDKGNIIAYDDGTDLEYTEKLDIISKI 120
DB 1394 DLKKLKKKFRVDDFTSEGRMEYDYKYDDKGNIIAYDDGTDLEYTEKLDIISKI 1453
QY 121 YGVLSPSKDGFEILGKISNVSKNAKYGGNNYSIRIKATKYPHSHKTMFPDLVYANIND 180
DB 1454 YGVLSPSKDGFEILGKISNVSKNAKYGGNNYSIRIKATKYPHSHKTMFPDLVYANIND 1513
QY 181 IYDGLAFAGDMRLFVKONDQKAEIKIRMPERIKETSEYEVYSSYGAVIELGEGDLSKN 240
DB 1514 IYDGLAFAGDMRLFVKONDQKAEIKIRMPERIKETSEYEVYSSYGAVIELGEGDLSKN 1573
QY 241 KPDNLITMNSGKIVSDSEKQOYLLKDNIIILKGYALKVTYNNPGKTMLEBNGVYSKEDI 300
DB 1574 KPDNLITMNSGKIVSDSEKQOYLLKDNIIILKGYALKVTYNNPGKTMLEBNGVYSKEDI 1633
QY 301 AKIOKAMPNIALSETTIVYASRVNEDGRSTQVLSALDGFNIIIRYOVFTFRKNDKGEA 360
DB 1634 AKIOKAMPNIALSETTIVYASRVNEDGRSTQVLSALDGFNIIIRYOVFTFRKNDKGEA 1693
QY 361 IDKGNLVTSSKLVLFKQDKKEYTGSDKFNVEAIKEDGSMFLPDTKRVNLSMDKNYFNP 420
DB 1694 IDKGNLVTSSKLVLFKQDKKEYTGSDKFNVEAIKEDGSMFLPDTKRVNLSMDKNYFNP 1753

```



```

Db      240 IECOLK-NISSNMLSGENIDINJENKSGYKEIISKF-----NEKFLDLK--NSR 288
Qy      144 NAKVYNNYKSIIEIKATKYDFHKTPTDLYANINDIYDGLAFAGDMFLFYDNDQKA 203
Db      289 NKEEYVNDKNDIE-----NENALKEKIKNLKD-----NIKVENLTKTN 331
Qy      204 EIKIR-----MPEKIKETKSEYRYV----- 223
Db      332 BERLQYLYKEKALEKKINLKEEBSYLRBEIGKEDNVNNFNKLEKEKELKILKSEB 391
Qy      224 -----SSGVNIELEGEDL--SKNKPDLTYMESKIYS-----DSEKQOYLK 265
Db      392 IETLSTSLKNEISVNEBEIVLNEKLDNINKSCDSYISSININIKTKEDIEKEINIK 451
Qy      266 DNIL-----RKGYALKVTTYNPKTDMLEGNGVSKED-----IAKIQK--AN 307
Db      452 ENILLNENLKNSRNISLKIISLNKCK-KLKEKNAAYSLRLEANNYMLSLBEKHYEYN 510
Qy      308 PNLRALSETTYIADSRNVEDGRSTQSVLMSALDGFNIIRYQVTPFRKNDGKALDKDGNL 367
Db      511 RSYKTLMEHYSGKVDNIKG-----CEVLDGIIVKKELETAMEIALGAL--SNV 560
Qy      368 VPDSSKLVLFKGDKEYTGEDKFNVEALKEKGSWLFIDTKPVNLSMDKNYENP--SKSN 424
Db      561 IIDE-----NKAKILINYLKCK-----SLGRATFLPITTIQGR 594
Qy      425 KIYVRNPEFLRKGISDKGPFNMLRNVESVDNYLLYGDLDHIDTDFNI--KLAVYDG 482
Db      595 KAKINN-----VTREDF--LGIASDLIDY--DVKEFNIIIDYVIGRTLVAKD- 637
Qy      483 DINDMGKDYKANGFPKATVDMQGNVYLQYGYSDLMNAKAVNHQPLVDNKPENVIDPK 542
Db      638 --MDSALAKLAKLNYSPFIYTLRGEV-----INPG 665
Qy      543 GNT-----SIEYADGKSVFNINDKKNNGFDGEIOQHITYNKEKYSFNDIKQIDKTLMI 599
Db      666 GSLTGSIRKRASSI--ISRRK--RIEB-----TKCELBEFTKN- 700
Qy      600 KIYVQDPAANTTYKEFLNKGTEVSLKPRHVTVTIIONGKMSSTVSEEDPLPYKKG 659
Db      701 -----TIEFPKMN-----ILENKKIKIT--LDBENLMI--KD 728
Qy      660 ELKGYQFDGMEISGFEGKDA-----GYVINLSMD--TFIKPYFKKIEEKEEENK 709
Db      729 EI-----YNNIEIETFKGKLAHAKEDTERLRSSINISBEIKLTKDKIKOLEENINSOK 784
Qy      710 PTFDVSKKKQNPQVNHSQLNESHKEDLOREHSHOKSDSTK---DYATVLDKNNISSK 765
Db      785 QLEBELKAKKO--LNNHDIKEC--EDFLQNEEENVKNIKDKLIEYKIEKALDEMLVSIK 839
Qy      766 -----STYNNPK 773
Db      840 KELYSMDNTININENK 857

RESULT 7
US-09-820-843A-109
; Sequence 109, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820, 843A
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patent version 3.0
; SEQ ID NO 109
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:

```

```

; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: 914493994
US-09-820-843A-109

Query Match      5.5%; Score 222; DB 10; Length 861;
Best Local Similarity 19.3%; Pred. No. 0.00037;
Matches 162; Conservative 136; Mismatches 324; Indels 216; Gaps 37;

Qy      2 LGELASRPFNLAG-----KGSLLKQDTGVSEHH--QENESIKESFTI 47
Db      154 LNNINGAVFKNLDNVCNPLTNMLYIDKEG--KHLTGKEHYNAASNEYHNNKNTNMY 211
Qy      48 DRNISTTRDPEKDL-----KLIKKKFRVDDFTSBGRMBEYDYKQDKGII 98
Db      212 NNNNSYNNNNFCNNNYNDNNYNNNSNNKMGKNTYERSILYK--KEHDVADYEYNNKNTIR 268
Qy      99 AYD-----DGTDEYETEKLEDEIKSKTYGVLSPEKDGHFBI--LQISNVSKNAKAVYGN 151
Db      269 KNDSEKTVNDPPLHY-----SKNNYDIFLGDIIKKAHNEKKGNN 310
Qy      152 NYSIRIKATKYDFHKTPTDLYANINDIYDG--LAFAGDMFLFYDNDQKALEIKIR 208
Db      311 KYNNM-----HDNNSN-----NSNNVNLNNNNNSNNYNNIFPDNDENL----- 351
Qy      209 MPEKIKETKSEYRYVSSGVNIELEG--DLSK--NKPDNLTYMESKIYSDSEKQOYL 264
Db      352 -----TKSNFAPMFKNNNNNNVNEVTDIIKYLNK-----NSQGHSDGKNNNNNN 396
Qy      265 KNTILRKGYALKVTTYNPKTDMLEGNGVSKEDIKIQANPR--LRLALSETTYIAD 321
Db      397 GNNII-----NNNSNNKNNI--FGKSRHYEVNMYINNNNNNNI--SNKNGEASFTVD 447
Qy      322 SRNVEDGRSTQSVLMSALDGFNIIRYQVTPFRKNDG--EALDKQNLVTDSSKLVLFKGD 380
Db      448 NINTNSGRBEKISNTYAR--LIMKOISMIKEKNGGLDVLEKQNT--FGFL 494
Qy      381 DKEYTGDKFVVEALKXDGSLFIDTKPVNLSMDKNYENPESKNKIYVRNPEFLY----- 435
Db      495 DNNYQNTGSSNN-----NSLRKNNM--KENDIYSEKASKRIMDIFR 533
Qy      436 -----RGLISDKGPFNMLRNVESVDNYLLYGDLDHIDTDFNIKLAVYDGDINDMGKND 491
Db      534 TLNNGVLVSQK-----SLTVAGSVLNNNNNNYNNNSNNRKYQNNNNNNNNNM----- 582
Qy      492 YKANGFPDKVTDMGANVYLQYGYSDLNK--AVGVHIOFLYDNVYKPEVNIIDPKGNTSIE 548
Db      583 -----NNNNNSNNNNNNNNNNNNYKNNHAKYHSMNVYTKLIF--INNYSNNDGNNNSNNS 633
Qy      549 YADGKSVFNINDKKNNGFDGEIOQHITYNKEKYSFNDIKQIDK----- 595
Db      634 NSNNNVEHYNNNNKKN--FKNNKINNTHNLPNNKNNNNNNNNYNNINNNLNMENPPPEL 691
Qy      596 TLNLIKIVQDPAAN--TYKBEFLNKGTEVSELKPRHVTVTIIONGKMSSTVSEED 651
Db      692 SPNNSDINKNNAQGNINITPIINSILRLD--NEVDVHNSISENIGNAK--VSNVLDLSLK 748
Qy      652 FILPYKGELEKGYQFDGMEISGFEGKDAQVNLSDDTIK-----PYFKKIEEKK 704
Db      749 SILKASKQGNNNNVIIP--KNFNNNNN-----NNNSKFLNYSQOYPSHQOQOQH 799
Qy      705 EENKPTFDVSKKQNPQVNHSQLNESHKEDLOREHSHOK---SDSTKDYATVLDK 759
Db      800 QOQOQOQ--QOQOQTLIQYINSLHNDPKKGGPKNKERIPMKYPRBDGTTNETMETVREK 855

RESULT 8
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chientli, Achar

```

```

; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuecong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; FILE REFERENCE: S1237/7019
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 06/272,930
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: Patent version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; US-10-087-464-10

Query Match      5.4%; Score 218; DB 14; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.0014;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy      23 KDTTG-VEHHQENBSIK-----EKSFTIDRNISTIRDPENKDKLKKRREYVD 75
Db      232 KDNVGMEDYIKKKKTIENINELBSKKTIDKKNMATEB-----KKKLYQAQY 283
Qy      76 FTSEGRMEERYDYKDKGNIAVDGTDLEYETBKLDITSKIYGVLSPEKQGHFEL 135
Db      284 DLSTYKQLEB-----AHMLISV-----LE--KRIDYTK-----KNEWIKEL 319
Qy      136 GKISNVSKAKYVGNKYKSIKIKATKYDFH-----SKTVPDLNINIDVGLAFA 188
Db      320 DKINIKKPPRPNSSNTPTLTDKQKKEBHEKEIKIAKTIKYNIDSLFTDPL----- 373
Qy      189 GDMRLFVKDNDOKKAEIKIMEPEKIKETKSEYP-----YVSSYGVN-----TELGE-GDLSK 239
Db      374 -ELAYTLABKN-KNIDIAKVKTESTEPENGVYPLSYNDINNALINELINSFGDIL- 430
Qy      240 NKPNULTMSESGKIYDSEKQYV--LKDNIIL-----RKGVALKVTTYNPKTKML 289
Db      431 -NPDYTKSPSKNTYTDERKKFPIBIEKIKIEKKKIESDKSYEDSKSLN-----DIT 485
Qy      290 EG-----NGVYSKE-----DIAKIOKA-----NPNRLASETTIYADSRNVEDGRTQSV 334
Db      486 KEYEKLAMRIYDSKKNNNIDLTNPEKMMKRSYVVEKLTHTNFPAASYENSQNLBKLTJK 545
Qy      335 LMSALDGF---NIT-----RYQVTFPKANDKGAIDQGNLVTDSSKLVLFGKDKRYT 385
Db      546 ALKTMEDYSLRNIIVEKEIKYKKNLISKIENIEITLVE--NIKQDEQLF-----EKIT 598
Qy      386 GEDKFNVAIKEDSGMLFIDTRPVNLSDKNVFNFSKSNKIYVRNPEFLRGIKSIDKGF 445
Db      599 KDNKPEDEKILEVSDIYAVQOKVLL--ANKIDELKKTQLTKNVE--LKNHIVPNSTY 653
Qy      446 NWEILRVNSVVDNYLYGDLMIDNTRDFNIXL-----NVK----- 480
Db      654 KOENKQOB---PYVLIYVAKKEIDKLKVFMPKVESLINEKKKIKITKEGSDNSEPSTGBEI 709
Qy      481 -----DGDIDMGKQYK----- 501
Db      710 TGOATTKRGGQAGSALDEDSVQAQAOEQKQAPVPVPVPEAKAVPTPPAPVNNKTRNV 769
Qy      502 TDMGNVYLTQGYSDINA-----KAVGVHYQFLYDNVPRVAVNIDKNGTSIEYADGKSV 556
Db      770 SKLD---YLEKLYBELANTSYICHKILYVSHSTMNKILKQYKITYEBSKLSGCDPLDL 826
Qy      557 FNIN-----DKRNGPFG---BIQCHITVNGKRTSFDNKOIIDFTLNIKIV 603
Db      827 FNIGNIPVNYSMFSLNNSLSQLEFMEIYEKEMVGNLYKLDKNDKIKMLLEBAKKVSTSV 886
Qy      604 KDPANVTVKEFLINKQGVSELSKPHRVTVTIOGKMSSTIYVEEDFPIIPVYGELEK 663
Db      887 KTLSSSQPLSLTPODKPEVSAANDTSHSTVNLNSLGFENILS-----LGKNKN 937
```

```

Qy      664 GYFDGMEISGPBK-----KQAGVYINLSKDTFIPVPEKIEKKKEBNKPTF--D 713
Db      938 IYQ-----ELIQKSSBNFYKILKSDPTFYNSSFTNFKVSKADNLSNDESKRKLLED 993
Qy      714 VSKKDNQPVNNSQLESHERKEDLQREBSQKSDSTKQVATVLDKNNISSK-STTNP 771
Db      994 INKLKTKQLSLSPDLNRYTKLBERLPDKKKYTGKTKMQIKKILKEQLSKLNSLNP 1052

RESULT 9
US-10-282-122A-70251
; Sequence 70251, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIYTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 70251
; LENGTH: 1141
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-10-282-122A-70251

Query Match      5.3%; Score 212; DB 15; Length 1141;
Best Local Similarity 20.9%; Pred. No. 0.0019;
Matches 182; Conservative 116; Mismatches 336; Indels 238; Gaps 44;

Qy      33 QENBSISGE--KSSFTIRANISTIRDPENKDKLKKR-----EYDDFTS 78
Db      221 KNNPEKALELVDNNTSTKPVATPTVAAPKLNKMPRAVAQAPAAVASNNVNDLIT 280
Qy      79 ETGRMEERYDYKDKGNIAVDGTDLEYETE-----KLDEIKSKIYGVLS 125
Db      281 VTKQYIK-----VGDGKNVAAAHGDKDIEYDEFTINDKVKRGGDMTINIDGNVPSDLT 336
Qy      126 PSKD-----GHEFELIGKISNVSKNAK---VYGNKYKSIKIKATKYDFHSKTMTDL 174
```



```

Qy 522 VGVHYQFLYDNVKEVNIDPKGNTSIEYADGKSVFNINDK-----RNGFDEIGROH 575
Db 624 IDKEDLMTLDS--HDPFTYDNDGNLEIH-----TOLANDINDDLKOKALNNANVKGIVVQ- 677
Qy 576 IYINGKEVTSNDIKQIIDKTLNIRKIYVCDPAPANTVKE--FILNKGGEVSELPKHVT 633
Db 678 ---DIEH---EIDVSIQANGKVIIPKMLANDPTENITTLK----- 716
Qy 634 VTIONGKEMSTIVSE-----DFTLPVYKGELEKGYOPGWEIS----- 673
Db 717 VLKQNNQPNIDLSIEEQLSGDNHISFKPKPTIAKTEN--DYEISFSPNSLANKKIK 773
Qy 674 -----GREGKQAGVYNLSQDTIKPVPKIKIEKKEBENKPTDV 714
Db 774 LFTPTDNNNTKTYEASIGLDGA-----IFKTSDAIPADPHKTYTLTKIEADNKKVANI 829
Qy 715 SKKQDPQVNHSSOLNESHREKDLQREHESQKSDSTQDVAATVLDKN----- 761
Db 830 DEISPLDRIVKQKNGNANADN--KHFKIPDQKNDLJAVYKDKNNNEIHVPIKTDDG 887
Qy 762 ---ISSKSTNNPNK 773
Db 888 KVIYVFNNNLFPDK 902

```

RESULT 11

```

US-10-282-122A-52048
: Sequence 52048, Application US/10282122A
: Publication No. US20040029129A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari
: APPLICANT: Zyckind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: EILTRA.034A
: CURRENT APPLICATION NUMBER: US/10/282,122A
: CURRENT FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/230,335
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/230,347
: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 78614
: SOFTWARE: Patent version 3.1
: SEQ ID NO 52048
: LENGTH: 1191
: TYPE: PRT
: ORGANISM: Clostridium acetobutylicum

```

US-10-282-122A-52048

Query Match 5.1%; Score 207; DB 15; Length 1191;
 Best Local Similarity 19.6%; Pred. No. 0.0038;
 Matches 178; Conservative 139; Mismatches 275; Indels 314; Gaps 43;

```

Qy 1 KUGELAEKPYKLNKNGKGG-----SLKQTYVHHHQBBSI 39
Db 120 RLKDIQ-LEFMDYTGKGEYSIIQGGKIAVLGSPREERRALIEBAGIVKFKTRVD- 177
Qy 40 KKSSEFTDRINISTRD-----FENKDLKLIKKEFREYD--FTSETG--RMEB 86
Db 178 -EKLEANTQNLVRINDLRTYERLEPLIESEAKFVELSDLKREINTIISIDN 236
Qy 87 YDYKDDKGNITIA---YDGT-----DLEYTEKLDI-----KSIYGVLS 126
Db 237 IDYRINDLKQMAADLKSIDENVKDEKISLELKVATESLDEFPDAKYSNKTXY- 292
Qy 127 SKDGHPELLGKI---SNVSNKAYYGNYSIE-----IKATKYDFHSKWTDPDY 175
Db 293 SKSEHQKLSELEIKERTSNSDVAKNKLVEIEIDLNSIVNLK--SRVEIQKLITED- 349
Qy 176 ANINDIVDGLAPAGDMRLFVKDNDQKAKIRIEMPE--KIKETYSB---YPVYSSYGN 228
Db 350 KNYNK-----ELSKINKSEKKNIDGLIEWEMKSIKQYKNDADIDISTISQNNN 400
Qy 229 VIELEBQDLSSKRPDNLTRMES---GKIYDSEKQOVLKDNITLRKGYALKTYTPNG 284
Db 401 EAVI---LKKIEISENESKLSIRKQAGYSNS-----LKINEYTKNTLSSELVKIN-- 448
Qy 285 KTDMLBAGVYSKEDIKIOKANPULRALSETTIYADSRNVEDGSTSVLMSALDGFNI 344
Db 449 --DKISYBQIRERKRSKISLN--RIISB----- 475
Qy 345 IRYQVFTKMDKGBAIDKQDNLVTDSKVLFGDKRYTGEDKFNVAIKEDGSMLFI 404
Db 476 --EKLANELNSKSKLEANKMLNL-----BKQYBQYNR--SVANLQHYTKGFV 522
Qy 405 DTKPVNLSMDKNYFNPSSNKIYVGNPEFYLRGKISDKGFPWELRVNBSYVD--NYLI 461
Db 523 DVKPESSVLEBVIKVKKEPFAV--SISLAALSD-----ITLDLDAKGLINYLK 573
Qy 462 YGDLDHNTDIPNI---KLNVKQD----- 482
Db 574 SKNLGRATFLPLNIIKGGKLNISDTRHEKFGIGIASLIDVDSFELPAVNVLAGRTVIV 633
Qy 483 DIMQKMDYRANGPDPVVTMDGNVYLQGYSDINAKAVGVHYQFLYDNVKEPVNIDPK 542
Db 634 DNMOSALKIATLNSYSFPIVTLTGEV-----VNG 663
Qy 543 GNTS--IEYADGKSVFNINDKNNNGFPDGBIEQHIEYINGKEVTSFNDIKQIIDKTLN 600
Db 664 GSLTGGSTYSKQASIT-----GRREIEBELNLENNNSQA---LEQSNKIIENK 710
Qy 601 IYVCDPAPANTVKEFILNKGGEVSELPKHRTVITIONGKEMSTIVSEDEPILPVYKGE 660
Db 711 KVVXG-----LDNLCLD-----LTDTI--HGKELKTKIKERLSIDIESEK 750
Qy 661 LEKYQFGMEISGEBGKQAGVYNLSKDTPIKVPFKIEK--KEBENKPTDVSK 717
Db 751 LKRSYNTSVGEG-PIEKIRIKHLEKLVKEBENNA--LKLR 789
Qy 718 KQNPVNHSSOLNESHRE-----EDLQREHESQKSD-----STQDVAATVLDK 759
Db 790 EAN---NNMLIDELERLKDERSKVLNBERIMSKVYKASDNLMSSTRIBRYKVM 846
Qy 760 NNISK 765
Db 847 HMEWK 852

```

RESULT 12

US-10-744-616-7

Sequence 7, Application US/10744616
Publication No. US20050026170A1
GENERAL INFORMATION:
APPLICANT: Patti, Joseph M.
APPLICANT: Foster, Timothy J.
APPLICANT: Hook, Magnus A.O.
APPLICANT: Bidlin, Delidre N.
APPLICANT: Petline, Samuel L.
TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from *Staphylococcus aureus*
FILE REFERENCE: P06283052/BAS
CURRENT FILING DATE: 2003-12-24
PRIORITY FILING DATE: 1997-11-26
PRIORITY FILING DATE: 1998-08-31
PRIORITY FILING DATE: 1998-08-31
NUMBER OF SEQ ID NOS: 23
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 7
LENGTH: 1166
TYPE: PRT
ORGANISM: *Staphylococcus aureus*
US-10-744-616-7

Query Match 5.0%; Score 203; DB 17; Length 1166;
Best local similarity 21.6%; Pred. No. 0.0062;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

33 QENESISE--KSFITIDENISTIRDPENKDKLKKKFR-----EYDDFTS 78
226 KNPPEKELVANDENTHSTPTVTAFTSVAPKVNAMKRAVAQAFAVANSNVNDLIK 285
79 ETGRMEYDYKYDKGNIIAYDGTDLLEYETE-----KLEDEISKIYGVLS 125
286 VTKQTIK---VGDEKDVNAAHADGKDIERYDTEFTIDNKKKKDTMTINYNKAVIPSDLT 341
126 PSKD-----GHEEILGKISNVSKNAK---VYGNKYKSIIRIKATKYDFHSK----- 168
342 DKNDPIDITDPSGEVIAGTTPKATKQIYTFDVIDKEDIKSLTLISYIDKKTVPNE 401
169 ---TWTF---DLYANIN-DIVDGLAFAGD---NELPYK-DNDQKAEIKIRM-PEKIK 214
402 TSLMTFTAGKETQNTVVDYQDPMVH-GDSNIQSIPLKDLBDEKQTEQIYVNPILKS 460
215 ETKSEY---PYVSSYGVIRIBGSD--LSRKKPNLTKM-----ESGKIYSDSKQ 261
461 AINIKVDIAGSQVDYGN-IKLNGSTIIDQNTETIKYKVNSDQQLPQSNRIYDFSQYED 519
262 YLLK-DNIIIRKGYALKVTVNPGKTDMLGNGVYSKEDIARIQANPILALSETTIVA 320
520 VTSQPDN---KKSFEENNVAITDFG-----DINSAYI-----IKVSKITPTS 558
321 DSR-NVEDGRSTQSVLMSALDGFNIIRYQVPTFRANDG-----E 359
559 DGEIADIAGTSKRT--DKYGYNAYAGSNFIVTSDNGGSGDYVPEBEKLYKIGDYWE 616
360 AIDKGNLVTSOSKVLVPGKDKKEYTGEDKFNVAIIXDGSMLPDTKPVNLSMDXNY-F 418
617 DVDDKG-----VQSTDSKERPMANVL-VTLTYPDGT-----TTSVRTDANGHEEF 660
419 NPSKSNKIYV--RNPEPYLRKGLS-----DKGPFNMLVNESVVDNYLYIGDLHD 468
661 GGLDGEFTYVAFETPTGLPTKVGNTTDEKDSNGSSVTVKINK-----DMSLD 712
469 -----NTRDFNLIKLVNKG--DINDMGKDYKANGFPDKVTDMDGAVTLQGYSL 517
713 TGFYKBPKNAGDYWEDTNDGIDANEPGIDVAVT---LDSYGVK-IGTTTTDA 766
518 NAKAVGVYQFLVDNKEPVNIDPKGNTSIEY---ADGKSVFN--INDKRANGDFGLQ 572
767 SGK-----YKF-----TDLN-NGNVTVEFTAGTTPYKATTTADKDSNGL----- 807
573 EOHVIYINKEYTSFENDIKQIIDKTLINIKI VVKDFPANT--TVKEPI---LNKDFGEVBSL 627

808 -----TTGVYIKADNMNTLD-----RGFTYTPKYSGLDYYWYNSNDGXQDST 851
628 KEFR-VTVIIONGK-BMSSTIVSEBDFILPVTKGHEIKGYQDPGMHISGFEKGDAGYVI 665
852 KQIKVTVTLQNEKEGVIGTGTIDEN-----GK-----YRFD----- 883
686 NLSKQTFIKVPFKIEKKEBENKPFVDVSKKGNQV-----NHSQ--LINESHREKDLQ 738
884 NUDSGKT-KVITEKPAGLTQTVNTTED-DKQADGGEVVTITDHDFTLNDGYFEEDTS 941
739 REHSQKSDSTDO 751
942 DSDSDSDSDSDSD 954

RESULT 13
US-10-032-585-7646
Sequence 7646, Application US/10032585
Publication No. US20030180953A1
GENERAL INFORMATION:
APPLICANT: Terry, Koemer D.
APPLICANT: Bo, Jlang
APPLICANT: Charles, Boone
APPLICANT: Howard, Busey
TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
FILE REFERENCE: 10182-005-999
CURRENT FILING DATE: 2001-12-20
PRIORITY FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 8000
SOFTWARE: Patent In version 3.1
SEQ ID NO 7646
LENGTH: 1881
TYPE: PRT
ORGANISM: *Candida albicans*
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (1881)
OTHER INFORMATION: X-any amino acid
US-10-032-585-7646

Query Match 5.0%; Score 203; DB 14; Length 1881;
Best local similarity 20.8%; Pred. No. 0.011;
Matches 172; Conservative 132; Mismatches 300; Indels 222; Gaps 39;

22 KQDTTGV---HHQENESIKKSSFTIDN-ISTIDPFRKOLKULIKKFRVDDF 76
1140 KQATDEIIFTKYTHITDLOEHAHKQKQSFESERNDIKSNLDEANKELSD-NREKLSNLEKE 1198
77 TSETGRMEYDYKYDDKGNIIAY--DQGTDLLEYETKLEDEIKSYGVLSPSKDGHPFI 134
1199 KTELANKKATQSEKISDLETSVAISDEKSKSLKHIDIKRKIKLEFLLKGBEFTMPFK 1258
135 LKISNVSNAAVYVYGNKYKSIIRIKATKYDFHSKMTFPLVYANINDIVDGLAFAGDMLP 194
1259 KEQLQVNDKCK-----ELBAC-----LKLTEKEREINDILIRLEAA----- 1297
135 VQNDQKAEIKIRMPEKIKETKSEHYVSSYGVNTEIGEGLSNKKPNLTKMESGKIY 254
1298 KSDHTEKRLSL---LBDTKSB---SEKNVTKLNE-----QLEKLR 1334
255 SDSERQOYLLKONIIIRKGYALKVTVNPGKTD-----LBGNGVYSKEDIARIQ 305
1335 GERBKE---VRD---IQGLAKATTDMEKIKTTLDKYLKBSKDLDEK---NBSVDTLKK 1385
306 ANPNLR--ALSETTIYVDSRVNVEDGRSTQSVLMSALDGFNI-----IRQVP- 350
1386 EVNENKKEISILEDKODITTKYKEIAAQLETKTSNLSJTWELKTELELKKVREMLE 1445
351 ---TFKANDG---BAIDKGNLVTSOSKVLVPGKDKKEYTGEDKFNVAIIXDGSML 402
1446 ATSELTKQDNNQSLTEIEKTKAALTSSK-----DLEVCAGNKSRLQ-----DSLK 1493

```

QY 403 FIDTKPVLSMDKXVFNPSKNIYVRNPEFY-----LRGKISDGKFNMLRVNESYVD 457
Db 1494 SVKSLKFNFKNTQOETTSLDEKHEKQBEIVTLQTEKDRISBEKERAMLSRSESTV- 1552
QY 458 NYLIYGDHLINDTDFNKLAVKOGDIDMDKMGKDYKANGFPDXKVTMDGNVYLTQYSDI, 517
Db 1553 -----IKENHS-----DKISLSL-----SKI 1567
QY 518 NAKAVGHYQFLYDNVKEPVNIDPKGNTSIEY-----ADGKSVFNINDKRNNGPDSGI 571
Db 1568 NS-----IKENHSKEITTHNEQKTSLKODIAKLSDQHSQAOTQLEDKKNQKELKA 1618
QY 572 Q-BOHVYNGKEYSF-----NDIKQIID--KTINKIYVKKOPARTVYKSF-----ILKND 620
Db 1619 SLEKH-----NTSATSIEBKNOIKELSTIKSLKTELTSGDALKKOSQREYVTKTKNSD 1675
QY 621 T-----GEVSEIKPHRVTVTLONGKMSSTVSEEDFILPVYKGLKGYOPDGEWISGF 675
Db 1676 TESKLERQLEEL--KVMSDLOTDADRKLKGIATERIAL-----KSELET----- 1717
QY 676 BGKODAGVYVLSKDTFLKVPYFKIIEKKSEENKTFPVSKKQNPQVNHQSQMSHRK- 734
Db 1718 --VNSGSLSTSEBALTKTV--KSLKEREKELQPLSGNKSKELEBYIQHSDISEKRL 1774
QY 735 EDLQREHSQKSDSTK-----DVTATVLDKNNISKSSTNNPNK 773
Db 1775 TDELKERTKOPDSDSKKGLTELENDUTST---KKELETKTQTSRK 1817

```

```

RESULT 14
US-09-742-096-3
; Sequence 3, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DROULHE, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-BRYTHOCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773USODIV
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: P. falciparum
US-09-742-096-3

```

```

Query Match 5.0%; Score 200.5; DB 9; Length 1786;
Best Local Similarity 20.5%; Pred. No. 0.015;
Matches 204; Conservative 160; Mismatches 329; Indels 301; Gaps 52;

QY 6 AES--KPKK--LGNKGGSLKQDT-----TGVENHNOEN-----SRIK 40
Db 868 AESVTFPSNILEIOENTITNDTIEKLEELHENVLSALENTQSBREKEVIDIIEVK 927
QY 41 EKSSFTI-----DRNISTIRD--PEN-----KOLKGLIKKKFRVDDFTS 78
Db 928 EVAVTTLLETVQAEKESKANTITTEFENLEENAVSNENVAENLEKLEMYTNYVLDKYE 987
QY 79 ET-----GRKM--EYDYK-----DDKG--NIAYDGTDLLET-----EKLD-- 114
Db 988 ETVIISGSLNENMDKAFSEIIPDNVKGIOENLLT---GMFRSIEISIVIOSEBEKVDLN 1044
QY 115 -EIKSKIYGVSPSKDGFELIGKISVSKAKV-----YGNNGYSIRIKATKQYFH 166
Db 1045 ENNVSSILNDIENKGG--LNLKLNISSTEGVQETVTEHVEQNVYDVDPVPAK----- 1097

```

```

QY 167 SKTMTFDLYANINDVDGLAPAGMRLPVKNDQKAEIKIRMPKIKETKSEPYVSSY 226
Db 1098 -----DQFLGILMAGEL-----XEMFNEEDVKSSSDVITVEIIEKDEPVQKEKET 1146
QY 227 GNVIELEGGD-----LSKRPDNLTFM-----ESGCIYSDEKQOVLKONILIRKGYALK 277
Db 1147 VSIIEBENIVDVLEBEKEDLTDKMDAVEESITISDSKERTSINDK---EKDVSLV 1203
QY 278 VTTNPG-----KTDMLEGNGVSK--EDIAKIQANPNL----- 310
Db 1204 VEEVDNDMDSSVEKYLEKNNBEELMKDAVEINDITSKLEIBTQELNEVRADLLIKDMK 1263
QY 311 -----RAISFTIYADSNVEDGR--STQSYLSALDGPNIIRYVFFPKMDKGRADIK 363
Db 1264 LKLEIKALSE-----DSKRIIDAKODTLEKYLEEHD-----ITTTLEDEVVELKQV 1309
QY 364 DGNLVTDSKLVLPKQDKKEYTGEDKFNVAIKEDGSMLEFIDTKPVNL-----SM 413
Db 1310 EEDKIBKYSDL-----KQLEBDLKE--VKEIKELSELLEBYKELKTIETDILBEKKKI 1362
QY 414 DKNYFNPSKNIYVRNDEPYLRGKISDGKGFNMLRVNESYVD-----NYLIYGD 465
Db 1363 EKDHKEKEEBEABEIKOLEADILKEVS-----SLEVEEKKLEBYHLEKEVEHIIISDA 1417
QY 466 HINDRDNINIKLVND--GDIM-----DMGMDYKANGFPDXKVTMDGNVYLTQYSDLN 518
Db 1418 HIKGLEBDLE--EVDNLKSLIDMLKGMELDMDKESLEBDVTYKGERV---ESLKQVL 1473
QY 519 AKAVGHYQFLYDN-----VXPEVNIDPKGNTSIEYADGKSVFVNINDK--RN 564
Db 1474 SSALGMDBEQKTRKKAQPKLEBVLKBEVBEBKKIT-----KKYVRFDIKQKBPD 1528
QY 565 NGPDGEIOPQHUYNGK-----EYTSFNDIKQIIDTLN-----IKIYVKDPRK--N 609
Db 1529 EIVEVEMQDEDIIEBVEEDIEBDIEBDEKEDIDEDIDGDKQDEVDLIYQEKERIEK 1588
QY 610 TTVEKFIINKDGE--VSEIKPH-----RYVTIONGKMSSTVSEED 651
Db 1589 VKAKKKLEKKEVBEVSGLKQHVDEVMKYQKIDGVDKESKALSKRDVYVNLKQND 1648
QY 652 FILPVYKELKGYOPDGMIEISGFEGKKDAGVYVNL-----SKDTFLKPVK 698
Db 1649 FFSKV--KTFVKKYKFAAPFISAVAA--FASYVGFPTFLSSCVTIASSTYLLSKVDK 1705
QY 699 KLEBKESKNP-----TQDV-----SKKQNPQVNHQSQMSHRKEDLQR 739
Db 1706 TINKNK--REFYSVFDFIFQNLKHYLQQMEKFSKERNNVIEVT--NKAERKGNVQV 1760
QY 740 BEHSQKSDTKQVATVLDKNNISKSSTNNPNK 773
Db 1761 TNKTEKT-----TKYDKNNKVPKRRQTSK 1786

```

```

RESULT 15
US-10-415-253-2
; Sequence 2, Application US/10415253
; Publication No. US20040067236A1
; GENERAL INFORMATION:
; APPLICANT: Drouilhe, Pierre
; TITLE OF INVENTION: Immunogenic Compositions Comprising
; TITLE OF INVENTION: Liver Stage Malarial Antigens
; FILE REFERENCE: B45250
; CURRENT APPLICATION NUMBER: US/10/415,253
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: PCT/EP01/12349
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: EP00203724.0
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:09:54 ; Search time 181 Seconds

(without alignments)
2186.946 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026
Sequence: 1 KLGIAESKFNKNGXKGS.....ATVLDKNNISSKSTNNPNK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	4026	100.0	2119	2	Q9AHT5 streptococc
2	4026	100.0	2140	2	Q97RY6 streptococc
3	3793.5	94.2	2144	2	Q94AM8 streptococc
4	3709.5	92.1	2144	2	Q8DOP7 streptococc
5	270.5	6.7	2649	2	Q7RAS7 streptococc
6	265.5	6.6	1642	2	Q81B84 plasmodium
7	264	6.6	3504	2	Q81L45 plasmodium
8	260	6.5	1850	2	Q7RGP2 plasmodium
9	259	6.4	2757	2	Q7RRR9 plasmodium
10	258.5	6.4	1811	2	Q7REH9 plasmodium
11	257.5	6.4	2661	2	Q7RMS4 plasmodium
12	257	6.4	1389	2	Q7RBP4 plasmodium
13	255	6.3	1777	2	Q813P4 plasmodium
14	253.5	6.3	1127	2	Q9YV76 melanoplus
15	252.5	6.3	2227	2	Q81121 plasmodium
16	252.5	6.3	2723	2	Q7ROB6 plasmodium
17	251.5	6.2	2849	2	Q81HY4 plasmodium
18	251.5	6.2	3381	2	Q812V4 plasmodium
19	251.5	6.2	3519	2	Q81E65 plasmodium
20	250	6.2	1474	2	Q81L02 plasmodium
21	250	6.2	1650	2	Q77328 plasmodium
22	247.5	6.1	5767	2	Q81525 plasmodium
23	246.5	6.1	2033	2	Q81M18 plasmodium
24	245	6.1	2269	2	Q81LA2 plasmodium
25	242.5	6.0	2664	2	Q7RBL0 plasmodium
26	242.5	6.0	3317	2	Q8EMP8 mycoplasma
27	241.5	6.0	1033	2	Q81B88 plasmodium
28	241.5	6.0	2694	2	Q7RJ11 plasmodium
29	241.5	6.0	3063	2	Q6IMC1 plasmodium
30	240.5	6.0	2740	2	Q7RFS2 plasmodium
31	240.5	6.0	4433	2	Q81J15 plasmodium

32	240	6.0	1455	2	Q81XG8 plasmodium
33	240	6.0	10061	2	Q81321 plasmodium
34	239.5	5.9	2569	2	Q81B98 plasmodium
35	239	5.9	2586	2	Q7PDT7 plasmodium
36	238	5.9	1081	2	Q8XIL2 clostridium
37	238	5.9	3322	2	Q81XLO plasmodium
38	238	5.9	3628	2	Q968Y0 plasmodium
39	238	5.9	3704	2	Q81XK8 plasmodium
40	237.5	5.9	1104	2	Q7RSO8 plasmodium
41	237.5	5.9	3535	2	Q81C29 plasmodium
42	237	5.9	2511	2	Q81L44 plasmodium
43	237	5.9	5229	2	Q7RTF4 plasmodium
44	236.5	5.9	1114	2	Q97242 plasmodium
45	236.5	5.9	1738	2	Q81A15 plasmodium

ALIGNMENTS

RESULT 1
Q9AHT5 PRELIMINARY; PRT; 2119 AA.
AC Q9AHT5: 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine protease (Fragment).
GN Name=prtA;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4;
RX MEDLINE=21115976; PubMed=1179332;
DOI=10.1128/IAI.69.3.1593-1598.2001;
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
Langermann S., Johnson S., Koenig S.;
RA use of a whole genome approach to identify vaccine molecules
affording protection against Streptococcus pneumoniae infection.*;
RT Infect. Immun. 69:1593-1598(2001).
RL -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
an amide bond (By similarity).
DR EMBL: AF291699; AKJ9159.1; -.
DR HSRP: P00782; 2SRT.
DR MEROPS: S08.064; -.
DR GO: GO:0009986; C:cell surface; IEA.
DR GO: GO:0005618; C:cell wall; IEA.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004289; F:subtilisin activity; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR010435; DUF1034.
DR InterPro: IPR01899; Gram_pos_anchor.
DR InterPro: IPR003137; PA.
DR InterPro: IPR00209; Pept_S8_S53.
DR InterPro: IPR010259; Prot_inh_S8A.
DR InterPro: IPR01680; WD40.
DR Pfam: PF06280; DUF1034; 1.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02225; PA; 1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR Pfam: PF05922; Subtilisin_N; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR TIGRPFAM: TIGR01167; lpxTG anchor; 1.
DR PROSITE: PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE: PS00137; SUBTILASE_HTS; UNKNOWN; 1.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN; 1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN; 1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER 1 1

SQ SEQUENCE 2119 AA; 238227 MW; 5179B7F6B960A6A CRC64;
 Query Match 100.0%; Score 4026; DB 2; Length 2119;
 Best Local Similarity 100.0%; Pred. No. 2e-161;
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLGEIASEKFKNLGNGKSGSLKQDTTGVVHHQENBESIKKKSFTIDNISTIRDFENK 60
 DB 1313 KLGEIASEKFKNLGNGKSGSLKQDTTGVVHHQENBESIKKKSFTIDNISTIRDFENK 1372
 QY 61 DLKLLIKKKFRVNDPTSTGKRMERYKYDDKNNIAYDGTDLVETKLDLTKSKX 120
 DB 1373 DLKLLIKKKFRVNDPTSTGKRMERYKYDDKNNIAYDGTDLVETKLDLTKSKX 1432
 QY 121 YGVLSPSKDGHPHLLGKISNVSNNKAVYGYNNYKSIKATKXDPFSKMTFIDLNYIND 180
 DB 1433 YGVLSPSKDGHPHLLGKISNVSNNKAVYGYNNYKSIKATKXDPFSKMTFIDLNYIND 1492
 QY 181 IVDGLAFAGDMRLPVKNDQKABIKIRMPKIKETKSEYPVSSYGVNIELGEGDLSKN 240
 DB 1493 IVDGLAFAGDMRLPVKNDQKABIKIRMPKIKETKSEYPVSSYGVNIELGEGDLSKN 1552
 QY 241 KPDNLTAKESGKIYDSKQOYLKDNITILKGYALKATTYTPGKTDLKNGVYSKEDI 300
 DB 1553 KPDNLTAKESGKIYDSKQOYLKDNITILKGYALKATTYTPGKTDLKNGVYSKEDI 1612
 QY 301 AKIQKAMPRLALSETTYYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPMNDKGA 360
 DB 1613 AKIQKAMPRLALSETTYYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPMNDKGA 1672
 QY 361 IDKQGNLTVDSSKVLFGKDYKXTGDEKFNVEAIKEDSMLPIDTKPNVLSMDKXYPNP 420
 DB 1673 IDKQGNLTVDSSKVLFGKDYKXTGDEKFNVEAIKEDSMLPIDTKPNVLSMDKXYPNP 1732
 QY 421 SKSNKIYVRNBEPTLRGKISDKGFWMLRVNVSVDNLTLYGDLHIDTRPDKLANK 480
 DB 1733 SKSNKIYVRNBEPTLRGKISDKGFWMLRVNVSVDNLTLYGDLHIDTRPDKLANK 1792
 QY 481 DGDIMDMGKQYKANGFPDKVYTDMDGNVLTQYGSDDLNAKAVGHYQPLVYNVKEVNID 540
 DB 1793 DGDIMDMGKQYKANGFPDKVYTDMDGNVLTQYGSDDLNAKAVGHYQPLVYNVKEVNID 1852
 QY 541 PKGNISIRYADGKSVYVNIINRRNNGFDGEIOBHHIYNGKETSFNIDIKOIIDTLNIX 600
 DB 1853 PKGNISIRYADGKSVYVNIINRRNNGFDGEIOBHHIYNGKETSFNIDIKOIIDTLNIX 1912
 QY 601 IVVKDPANNTYKCEIILNKDNGEYSELKPHRYVTYTIQNGKMSSTIVSEEDFILPVYKGE 660
 DB 1913 IVVKDPANNTYKCEIILNKDNGEYSELKPHRYVTYTIQNGKMSSTIVSEEDFILPVYKGE 1972
 QY 661 LEKGYQFDGWEISGEGKKDAGVYINLSKDTFIKPVFKKIBEKKEBNKPTFDVSKKCN 720
 DB 1973 LEKGYQFDGWEISGEGKKDAGVYINLSKDTFIKPVFKKIBEKKEBNKPTFDVSKKCN 2032
 QY 721 POWNHQSUNESHRKEDLOREHSQKSDSTKQVATATVLDKNISSTSTNNPK 773
 DB 2033 POWNHQSUNESHRKEDLOREHSQKSDSTKQVATATVLDKNISSTSTNNPK 2085
 RESULT 2
 ID 097RY6 PRELIMINARY; PRT; 2140 AA.
 AC 097RY6;
 DT 01-OCT-2001 (TRENBLrel. 18, Created)
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
 DE Serine protease, subtilase family.
 GN OrderedLocustNames=SP0641;
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC BAA-334 / TIGR4;
 RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
 RA Tettelein H., Nelson K.E., Paulsen I.T., Eilen J.A., Read T.D.,
 RA Peterson S.N., Heidelberg J.F., Deboy R.T., Hatt D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
 RA Holtzapple E.K., Khouli H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
 RA Hickey B.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.,
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae."
 RL Science 293:498-506(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (by similarity).
 DR EMBL: AE007373; AKT74791.1; -.
 DR PIR: F95074; F95074.
 DR HSSP: P00782; 2SBR.
 DR MEROPS: S08.064; -.
 DR TIGR: SP0641; -.
 DR GO: GO:0009986; C:cell surface; IEA.
 DR GO: GO:0005618; C:cell wall; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0004289; F:subtilase activity; IEA.
 DR GO: GO:000508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR010435; DUF1034.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR000209; Pept_S6_S53.
 DR InterPro: IPR010259; Prot_inh_S8A.
 DR Pfam: PF06280; DUF1034; 1.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR Pfam: PF02225; PA; 1.
 DR Pfam: PF00082; Peptidase_S6; 1.
 DR Pfam: PF05922; Subtilisin_N; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR TIGRFAM: TIGR01167; LPTXG_anchor; 1.
 DR PROSITE: PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN 1.
 DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN 1.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN 1.
 KW Cell wall; Complete proteome; peptidoglycan-anchor; Protease.
 SQ SEQUENCE 2140 AA; 240426 MW; PA44AD8E2938B314 CRC64;
 Query Match 100.0%; Score 4026; DB 2; Length 2140;
 Best Local Similarity 100.0%; Pred. No. 2e-161;
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KLGEIASEKFKNLGNGKSGSLKQDTTGVVHHQENBESIKKKSFTIDNISTIRDFENK 60
 DB 1313 KLGEIASEKFKNLGNGKSGSLKQDTTGVVHHQENBESIKKKSFTIDNISTIRDFENK 1393
 QY 61 DLKLLIKKKFRVNDPTSTGKRMERYKYDDKNNIAYDGTDLVETKLDLTKSKX 120
 DB 1394 DLKLLIKKKFRVNDPTSTGKRMERYKYDDKNNIAYDGTDLVETKLDLTKSKX 1453
 QY 121 YGVLSPSKDGHPHLLGKISNVSNNKAVYGYNNYKSIKATKXDPFSKMTFIDLNYIND 180
 DB 1454 YGVLSPSKDGHPHLLGKISNVSNNKAVYGYNNYKSIKATKXDPFSKMTFIDLNYIND 1513
 QY 181 IVDGLAFAGDMRLPVKNDQKABIKIRMPKIKETKSEYPVSSYGVNIELGEGDLSKN 240
 DB 1514 IVDGLAFAGDMRLPVKNDQKABIKIRMPKIKETKSEYPVSSYGVNIELGEGDLSKN 1573
 QY 241 KPDNLTAKESGKIYDSKQOYLKDNITILKGYALKATTYTPGKTDLKNGVYSKEDI 300
 DB 1574 KPDNLTAKESGKIYDSKQOYLKDNITILKGYALKATTYTPGKTDLKNGVYSKEDI 1633
 QY 301 AKIQKAMPRLALSETTYYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPMNDKGA 360
 DB 1634 AKIQKAMPRLALSETTYYADSRNVEDGRSTQSVLMSALDGNFIIRYQVTFPMNDKGA 1693

```

QY 361 IDKGNLVTDSKVLFGKDDKEYTGEDKFNVAIKEDGSMLEFIDTRKPVNLSPMDKNYFNP 420
DB 1694 IDKGNLVTDSKVLFGKDDKEYTGEDKFNVAIKEDGSMLEFIDTRKPVNLSPMDKNYFNP 1753
QY 421 SKSNKLIYRNPEFLRGKISDKGFGNWEIARVNESVVDNYLYGDLHIDNTRDFNKLANK 480
DB 1754 SKSNKLIYRNPEFLRGKISDKGFGNWEIARVNESVVDNYLYGDLHIDNTRDFNKLANK 1813
QY 481 DGDIMDMGKDYKANGFPDXTTDMGNYVLTQGYSDLNAAKAVGVHYQFLYDNVKEPVNID 540
DB 1814 DGDIMDMGKDYKANGFPDXTTDMGNYVLTQGYSDLNAAKAVGVHYQFLYDNVKEPVNID 1873
QY 541 PKGNSTIERYADGKSVVFNINDRNNGPDGEIOBOHIYINGKRYTSPNDIKOIIDTLNKK 600
DB 1874 PKGNSTIERYADGKSVVFNINDRNNGPDGEIOBOHIYINGKRYTSPNDIKOIIDTLNKK 1933
QY 601 IVVADPARNTTVKEEILNKDTGEVSELKPHRYTVITIQNGKMSSTIVSEEDPILPVYKGE 660
DB 1934 IVVADPARNTTVKEEILNKDTGEVSELKPHRYTVITIQNGKMSSTIVSEEDPILPVYKGE 1993
QY 661 LEKGYQPDGWEISGEGKKDAGYVNLSDPTFKVFKIEKKERENKPTFDVSKKDN 720
DB 1994 LEKGYQPDGWEISGEGKKDAGYVNLSDPTFKVFKIEKKERENKPTFDVSKKDN 2053
QY 721 POWNHOLNESHKREDLQREBHSOKSDSTKDYATATVLDKNNISKSSTTNPNK 773
DB 2054 POWNHOLNESHKREDLQREBHSOKSDSTKDYATATVLDKNNISKSSTTNPNK 2106

```

RESULT 3

```

ID 09S4M8 PRELIMINARY; PRT: 2144 AA.
AC 09S4M8
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase precursor Prta.
GN Name:Prta.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3 B:
RX MEDLINE=21585565; PubMed=11728722.
RA Behe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,
  Zyk G.;
RT "The cell wall-associated serine protease Prta: a highly conserved
  virulence factor of Streptococcus pneumoniae.";
RL FEBS Microbiol. Lett. 205:99-104(2001).
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
  an amide bond (by similarity).
DR EMBL; AF127143; AAD48399.1; -.
DR HSSP; P00782; 2SPT.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.

```

```

DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LepTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN; 1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN; 1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN; 1.
DR Cell wall; Peptidoglycan-anchor; signal.
FT SIGNAL 1
FT CHAIN 20 2144
FT PRTA. cell wall-associated serine proteinase
SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;
Query Match 94.2%; Score 3793.5; DB 2; Length 2144;
Best Local Similarity 94.7%; Pred. No. 1.3e-151;
Matches 732; Conservative 16; Mismatches 24; Indels 1; Gaps 1;
QY 1 KKGIAESKPKNLAGKESGLKQDTTGVHHHQBESIKESSTTIDNISTINDPNK 60
DB 1339 KKGIEPSKPKNLAGKESGLKQDTTGVHHHQBESIKESSTTIDNISTINDPNK 1398
QY 61 DKKLIIKKKRFREVDFTSPTGKMEYDYKYDDKGNITLAYDGTDLRYETKLDDEIKSI 120
DB 1399 DKKLIIKKKRYQEDFPVNGGRTYERDITDKGNITLAYDGTDLRYETKLDDEIKSI 1457
QY 121 YGVLSPSKDGHEILIGKISNYSKNAKYVYGNYSKIEIKATKYDFHSKMTFEDLYANIND 180
DB 1458 YGVLSPSKDGHEILIGKISNYSKNAKYVYGNYSKIEIKATKYDFHSKMTFEDLYANIND 1517
QY 181 IYDGLAFAGDMRLFYKNDQKKAELIKIMPEKIKETKSEYPVYSSGYNIEIAGEBLSKN 240
DB 1518 IYDGLAFAGDMRLFYKNDQKKAELIKIMPEKIKETKSEYPVYSSGYNIEIAGEBLSKN 1577
QY 241 KPDNLTKHESGKIYDSEKQOYTLKDNITLRGYALKVTPYKPTDMLBKGVSXKDI 300
DB 1578 KPDNLTKHESGKIYDSEKQOYTLKDNITLRGYALKVTPYKPTDMLBKGVSXKDI 1637
QY 301 AKIQANPNLRALSETTIVYADSRNVEDGRTSGVLSALDGFNIIRYQVFTFRMNDKGEA 360
DB 1638 AKIQANPNLRALSETTIVYADSRNVEDGRTSGVLSALDGFNIIRYQVFTFRMNDKGEA 1697
QY 361 IDKGNLVTDSKVLFGKDDKEYTGEDKFNVAIKEDGSMLEFIDTRKPVNLSPMDKNYFNP 420
DB 1698 IDKGNLVTDSKVLFGKDDKEYTGEDKFNVAIKEDGSMLEFIDTRKPVNLSPMDKNYFNP 1757
QY 421 SKSNKLIYRNPEFLRGKISDKGFGNWEIARVNESVVDNYLYGDLHIDNTRDFNKLANK 480
DB 1758 SKSNKLIYRNPEFLRGKISDKGFGNWEIARVNESVVDNYLYGDLHIDNTRDFNKLANK 1817
QY 481 DGDIMDMGKDYKANGFPDXTTDMGNYVLTQGYSDLNAAKAVGVHYQFLYDNVKEPVNID 540
DB 1818 DGDIMDMGKDYKANGFPDXTTDMGNYVLTQGYSDLNAAKAVGVHYQFLYDNVKEPVNID 1877
QY 541 PKGNSTIERYADGKSVVFNINDRNNGPDGEIOBOHIYINGKRYTSPNDIKOIIDTLNKK 600
DB 1878 PKGNSTIERYADGKSVVFNINDRNNGPDGEIOBOHIYINGKRYTSPNDIKOIIDTLNKK 1937
QY 601 IVVADPARNTTVKEEILNKDTGEVSELKPHRYTVITIQNGKMSSTIVSEEDPILPVYKGE 660
DB 1938 IVVADPARNTTVKEEILNKDTGEVSELKPHRYTVITIQNGKMSSTIVSEEDPILPVYKGE 1997
QY 661 LEKGYQPDGWEISGEGKKDAGYVNLSDPTFKVFKIEKKERENKPTFDVSKKDN 720
DB 1998 LEKGYQPDGWEISGEGKKDAGYVNLSDPTFKVFKIEKKERENKPTFDVSKKDN 2057
QY 721 POWNHOLNESHKREDLQREBHSOKSDSTKDYATATVLDKNNISKSSTTNPNK 773
DB 2058 POWNHOLNESHKREDLQREBHSOKSDSTKDYATATVLDKNNISKSSTTNPNK 2110

```

RESULT 4

```

ID 08DOP7 PRELIMINARY; PRT: 2144 AA.
AC 08DOP7;

```

DT 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 DE Cell wall-associated serine proteinase PrtA (BC 3.4.21.-).
 GN Name=prta; OrderedLocustNames=sp0561;
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus;
 NC NCB1_Taxid=171101;
 RX DO=10.1128/JB.183.19.5709-5717.2001;
 RA Hocking J., Alborn W.B. Jr., Arnold J., Blaszcak L.C., Burgett S.C., Dehoff B.S., Bacterem S.T., Fritz L., Pu D.-J., Puller W., Geringer C., Gilmour R., Glast J.S., Knoch H., Kraft A.R., Lagace R.B., Leblanc D.J., Lee L.N., Letkowitz B.J., Lu J., Matsushima P., McLearen S.M., McInerney M., McLeaster K., Mundy C.W., Nicas T.I., Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P., Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G., Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R. Jr., Skatrud P.L., Glas J.I.;
 RA "Genome of the bacterium Streptococcus pneumoniae strain R6.";
 RL J. Bacteriol. 183:5709-5717(2001).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by an amide bond (By similarity).
 DR EMBL: AB008434; AK9365.1; -.
 DR PIR: A97942; A97942.
 DR HSSP: P00782; 2S8T.
 DR MEROPS: S08.064; -.
 DR GO: GO:0009986; C:cell surface; IEA.
 DR GO: GO:0008233; F:peptidase activity; IEA.
 DR GO: GO:0004289; F:subtilase activity; IEA.
 DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro: IPR010435; DUF1034.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR003137; PA.
 DR InterPro: IPR000209; Pept_S8_S53.
 DR InterPro: IPR010259; Prot_inh_S8A.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF06280; DUF1034; 1.
 DR Pfam: PF00746; Gram_pos_anchor; 1.
 DR Pfam: PF02225; PA; 1.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR Pfam: PF05922; Subtilisin_N; 1.
 DR PRINTS: PR00723; SUBTILISIN.
 DR TIGRPFAM: TIGR0167; LPTG_anchor; 1.
 DR PROSITE: PS0847; GRAM_POS_ANCHORING; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN; 1.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN; 1.
 KM Cell wall, Complete proteome; Peptidoglycan-anchor.
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC03A0C CRC64;

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
 Best Local Similarity 92.4%; Pred. No. 4.5e-148;
 Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;

QY 1 KLGRIAEKFKNGKXGSLKQDTGVHHQHEESIKESSTIDRNSTIDFNK 60
 DB KLGRIAEKFKNGKXGSLKQDTGVHHQHEESIKESSTIDRNSTIDFNK 1397
 QY 61 DLKGLIKKKFRVDDFTSETGRMEYDYKYDDKNIIAYDGTDLLEYTEKLDIKSKI 120
 DB DLKGLIKKKFRVDDFTSETGRMEYDYKYDDKNIIAYDGTDLLEYTEKLDIKSKI 1457
 QY 121 YGUSLSPGDGHPILIGKISNYSNAKYKANNYSIEIKATYDTHSTMTFDLYANND 180
 DB YGUSLSPGDGHPILIGKISNYSNAKYKANNYSIEIKATYDTHSTMTFDLYANND 1517
 QY 181 IVDGLAFAGDRLFLVYKNDQKAKIKIMPEKIKETKSEYFVSSYGVNIELGEGDLSKN 240
 DB IVDGLAFAGDRLFLVYKNDQKAKIKIMPEKIKETKSEYFVSSYGVNIELGEGDLSKN 1577

QY 241 KPDNLTMSGKTVSDSEKQVTLKDNITLRGVALKVTYTPRGKTDMEGNGYSKEDI 300
 DB KPDNLTMSGKTVSDSEKQVTLKDNITLRGVALKVTYTPRGKTDMEGNGYSKEDI 1637
 QY 301 AKIQANPLRALSETTIIYADSRNVEDGRSTQVMSALDGFNIIIRYQVTFEAMDNGBA 360
 DB AKIQANPLRALSETTIIYADSRNVEDGRSTQVMSALDGFNIIIRYQVTFEAMDNGBA 1697
 QY 361 IDKGNLVTSSKLVLPQGDKEYTGEDKFNVEAIKEDGSMFPIOTKPVNLSMDKNYFNP 420
 DB IDKGNLVTSSKLVLPQGDKEYTGEDKFNVEAIKEDGSMFPIOTKPVNLSMDKNYFNP 1757
 QY 421 SKSNITVYRNPEFYLRGKISDCKGNMELRVNESVVDNLYIGDILINDTRDFNKLANK 480
 DB SKSNITVYRNPEFYLRGKISDCKGNMELRVNESVVDNLYIGDILINDTRDFNKLANK 1817
 QY 481 DGDINDWGMKDKRANGFPDVKVTMDGNVYLQTVSDNLNAKAVGVHQPFLYDNVKEVND 540
 DB DGDINDWGMKDKRANGFPDVKVTMDGNVYLQTVSDNLNAKAVGVHQPFLYDNVKEVND 1877
 QY 541 PKGNTSIEYADGKSVYRNINDKRNNGPGEIIOBHIIYNGKXTSPNDIKOITDKLANK 600
 DB PKGNTSIEYADGKSVYRNINDKRNNGPGEIIOBHIIYNGKXTSPNDIKOITDKLANK 1937
 QY 601 IYVVDPAFNTVKEFTLNKDGSEVSELKPHRYVTYIIONGKEMSTIYSEDFILPYVKG 660
 DB IYVVDPAFNTVKEFTLNKDGSEVSELKPHRYVTYIIONGKEMSTIYSEDFILPYVKG 1997
 QY 661 LEKGYPDGMELSGEGKKDAGVYVNLSDPTFKPCKIEBKKEBNKPTFVDSKKDN 720
 DB LEKGYPDGMELSGEGKKDAGVYVNLSDPTFKPCKIEBKKEBNKPTFVDSKKDN 2057
 QY 721 PQVNSQINESHRKEDLQREHSQKSDSTKVYATVLRNNISKSTTNPNK 773
 DB PQVNSQINESHRKEDLQREHSQKSDSTKVYATVLRNNISKSTTNPNK 2110

RESULT 5
 QY 07RAS7 PRELIMINARY; PRT; 2649 AA.
 AC 07RAS7;
 DT 01-MAR-2004 (TREMblrel. 26, Created)
 DT 01-MAR-2004 (TREMblrel. 26, Last sequence update)
 DE 01-MAR-2004 (TREMblrel. 26, Last annotation update)
 GN Name=py06422;
 OS Plasmodium yoelii yoelii.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NC NCB1_Taxid=73239;
 RX PDB: 1YXN;
 RX PubMed=12368865; DOI=10.1038/nature01099;
 RA Carlton J.M., Angluoli S.V., Suh B.B., Koell T.W., Pierce M., Silva J.C., Ermolaeva M.D., Allen J.B., Selengut J.D., Koo H.L., Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell S.L., Shallow S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V., Cho J.K., Quackenbush J., Sedegah M., Shoabdi A., Cummings L.M., Plotens L., Yates F.R. III, Raine J.D., Sinden K.E., Harris M.A., Cummings D.A., Preiser P.R., Bergman L.W., Valdivia A.B., van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R., Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J., Garucci D.J.;
 RA "Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii.";
 RL Nature 419:512-519(2002).
 CC -1- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is preliminary data.
 CC EMBL: AAB01002172; EAA18637.1; -.
 DR InterPro: IPR011591; Botulinum.
 DR ProDom: PD001963; Botulinum; 2.

KW Hypothetical protein.
 SQ SEQUENCE 2649 AA; 309810 MW; E7207F344643AC24 CRC64;

Query Match 6.7%; Score 270.5; DB 2; Length 2649;
 Best Local Similarity 20.6%; Pred. No. 0.0052;
 Matches 206; Conservative 160; Mismatches 327; Indels 307; Gaps 49;

1 KLGEIAESKPKNLGN-----GKESGLKDDTGVGHHHGHQENBESIKESKSPFIDRNISTID 56
 1044 KNNLIERNYRNDNLHGVNKAKKNNKIDSNLILNKSFGDNYVQNRSYITTFETFTIS 1103
 57 FE-----NKDKKLIKKEFREVDFTSETGKMEBYDYKDDKNIIVAYD----- 102
 1104 FEDKIKILQNERLKKLEKLEKDEKGRKNDY-KKMEKIKKQD--NLFEATDGHIEK 1159
 103 -GTDLEETKDELKSKIVGVLSPSKGFELIGLISVNSKRAKAYYGN--NYKSIHK 159
 1160 LHCELENNKSKONELKINL-----KDKMKKILILESQICNN-NVNSHEHLNHTLMDVK 1211
 160 ATKVPFSKWTFFDYANINDIVDGLAFAGDM-----RLFVKNDQCKAE-IKIRMPK 213
 1212 NSENNYKHTNNNDVHSLNDSITDKKKIKKQIVLDQQLPALKNEIKQMEFLSKVVELN 1271
 214 KETKSEYVSSYGVNVELSGDLSKPKDNL----- 245
 1272 KLNKKKNNIMEYERSIDKLEKSIDKLEKQNLKLEQICDLKKNIMEKAAQIRDESS 1331
 246 -----TQESG-----KLYSD----- 256
 1332 NTTSSIDSGTTINNEIKIMEKIEALYKDKIKLKSNLEKTKNINILNTLKTSNBS 1391
 257 ---SEKQOYLKDNILRKGVALKYTYNPKGTMLENGVYSKEDIQKAPMLRAL 313
 1392 IELNKKIKLAKENKQLQNKYKSI-----NDLKANIKVD-EDILKENSFKPSII- 1442
 314 SETTYADSRNVEDGKSTQVLSALD--GNIIIR-YQVTFKRNKG-----EALD 362
 1443 --TFEDLNNEKIEHNTVQTDSTQIDDSYDNRKIKGGYDISHKONSRCYQDIKTEID 1500
 363 KD-----GNL-----VTDSKLVFGKDKKEYT--GDKFNEALIKESQMLPIDTKPV- 409
 1501 KMMKEKCTPMLKETOQDIDTLVLSVDKKEKTEKQDDYN--IIEEDNLKINENSNF 1557
 410 NLSMDKNYFNPSKSNKIYVRNPEFY-----LRKISDKG--FWMELRVNESVVD 457
 1558 NNRKENITYNLKNENKTVNANNIFEIDIRANLQWMPFNANGSNFNNYEVANNNE--H 1614
 458 NYL-----YGDLDIDTRDPNFK--LNVKQGDIMGMKDYKANGPDKTYTMDGNV 508
 1615 NYTNMKNKIMNAYSNSHLSN--NISSEHVSQNTSILAQ-----NYNKKYVNTKIYADIBV 1664
 509 YLQGTGYSDLNKAVGVH-----YQFLVDNPKPVNID 541
 1665 YNRN-----NLQFISLHNDENNRRTISNGTISNVSERKONSNTKYVANNNSKEITEN 1719
 542 KGNTSIEYADGKSVFVNIDK-----RNNGPDGELIQEHIYINGKEYTSFNDIKQIDK 595
 1720 NNNNSIKNSD-----NVNKKYECIKINBESANNESTLNTTENNSTNDLKNIYIE 1773
 596 TLNFKI-----VVDPAKNTTVKFIIL-KOTGVSSELKPRVYVITLONGEMSTIYS-EE 650
 1774 DNNLIAVNNKIIENY-NDQDLKNYILNSQKTNMNNHKEPND-NMTVEKKGKGEKGMIIIDIN 1831
 651 DFLIPVYKGELEKGYQFDGMBISGFBGKAGVILSKOTFLKPVFKIKBEKKEENRP 710
 1832 NEVLPYTK--IEKCVLSD-----EKBACTKKNKK--KSCQKNSKNSKT 1872
 711 -----TFDVSX-----KQDNPOVN-H-----SQLESER 733
 1873 VNNMNNRTYNIYKRPSEBSIKNKTEPANKNNPILMKTNGKLDLVLNDISKLYKNCI 1932
 734 KEDLQREHSQKSDSTKQVT--ATVLDKNNISKSTTN 770

DB 1933 KEELKGNIS-----KQITNLEITTIKKKNPKISKSLAN 1966

RESULT 6
 ID 081B84 PRELIMINARY; PRT; 1642 AA.
 AC 081B84;
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE Hypothetical protein MAL8pl.29.
 GN Name=Mal8pl.29;
 OS Plasmodium falciparum (isolate 3D7).
 OC Baktayota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_Taxid=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Seeger K., Murphy L., Hattie D., Berriman M., Pain A., Hall N.,
 RA Quail M., Barrell B.,
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL844507; CAD51123.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 1642 AA; 194753 MW; 701F5D28B2BE8097 CRC64;

Query Match 6.6%; Score 265.5; DB 2; Length 1642;
 Best Local Similarity 20.4%; Pred. No. 0.0048;
 Matches 190; Conservative 133; Mismatches 313; Indels 297; Gaps 46;

4 ELAESKPKNLGNKESGLKDDTGVGHHHGHQENBESIKESKSPFIDRNISTID 60
 104 EINRNNNNRNNRNNNSNSNPFIDBKEKTYNRTIMLKHKKRYDPDNTS-----DEK 158
 61 DLKLIKKEKREVDFTSENGKRM-----EEDYKDDKNIIVAYDGTDLLEYETKLD 115
 159 NIKR-----KKINDINIKKKKKMILPSYBENDKSNKSPFISD-----DE 201
 116 ISKIVGVLSPSKGFELIGLISVNSKRAKAYYGNVYSIEIKATKYDFHSKTMFDLY 175
 202 M-----DNKSNVNISSGK-----NY-STENNNDKNNPHISL--EKM 235
 176 ANINDIVDGLAFAGDMRLVYKNDQCKKAIKIRMPKIKETSEYVSSYGVNVELSG 235
 236 DKINN-----EKKEY-----STKSDIDLRVSIHSGYE-----NNNIDKLGCH 275
 236 DLKKNPDMLTWESKSIYDSEKQOYLKDNILRKGYA----- 275
 276 NLRNNGSNLVNPDNRK-----RKENPFIHANNMKKCKENKNTNINSYDIENHDEKDK 330
 276 -----LKVTTNPGK--TMDLGN-----GVYSKG-DIAKIOKANPMLRALS 314
 331 NEILNNLIDINSYVERKCLSYNNKRENDPFKIGISKOPNVDSKERYKYEKDHNIYNDLS 390
 315 ETTIYADSRNVEDGR-STQVLSALDGNFIIRYQVTFPMNDK----- 357
 391 DQKSGIDRTNVYDNNSSSHVSLSNL--FNESNFMLEDKKIRDKYKGVKLIKSLERNNS 448
 358 -----GEALDXD--GNLYTD-----SSRLVLFQGDKEYTGBDKFNVNAIKS 397
 449 ELYENELKEGETHKOBLKVLVDMTITKOINEKKKVVILYHKQYS-----NVERKSIS 503
 398 DQSGMLFIDTPVPLNSMDKNYFNPSKSNKIYVRNPEFYLAGKISDKGKFMELAVNESVVD 457
 504 EN-----IDIGPTNRSFLVGNNTKQMTHTINENNDR-----DKNNND--INN--D 544
 458 NYLIVGDLHIDTRDPNFKLNVYDGDIMGMKDYKANGPDKTYTMDGNV-----YL 510
 545 NYIINNNDNINNNNYVNVK--ESNEDIF-----YANGLSKGIDNIYQNIIRIKKIKS 536
 511 QGTGYSDLNKAVGVHQLFYDNV-KPEVNIIDPKGN--TSIEYADGKSVFVNIDKRN 565
 597 NSGNVYIDAN-----LSDGISNDEANLKVNTNSMDYTKLYDDDK-----GENTK 642
 566 GPDGELIQEHIYINGKEYTSFNDIKQIDKTNINIKI VVNDPAKNTTVKPIILKQIGVS 625


```

Db      643 TF-----KSAVDXKXKLLIENDTLMHNNNN-----QKEMIND--EKTLNGINNSPK 691
Qy      626 ELKPRVYTTIO---NGKEMSTIVSEDPFLPVYKGLBKGYQPDGWEISGPEGKD- 680
Db      692 DLDRKSNIDIOVNNNNNNNNNDTTFYDAMEYSIGNRBOYTRKVOBD-HVIMGVESKNN 750
Qy      681 -----AGVNLKSDTFL-----KPYE-- 697
Db      751 MGTNCSNNKRYINNDRNHLKEXYFDMMKPGVVDNIINIKENELVQNRKKKDIIFNK 810
Qy      698 ---KXIEKKEBENT-PTFDVSKKDPQVNH-----SQLNESH--KEDLQRE 740
Db      811 SHGNEKIEPLKOPNNKGVKVTNLLKVKVNDIDHEDVSVNIGRIDENRKKKKENILNS 870
Qy      741 EHSQSDSTKD-VTATVLDKNNISSEKSTNNPN 772
Db      871 INQEKDIGNNKINIKTSYSKNNILSKSYAKEPH 903

```

RESULT 7

```

ID      081145      PRELIMINARY;      PRT;      3504 AA.
AC      081145;
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      Hypothetical protein.
GN      ORFName=PF14_0404;
OS      Plasmodium falciparum (isolate 3D7).
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=36329;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=2255705; PubMed=12368864; DOI=10.1038/nature01097;
RA      Gardner M.J., Hall N., Rung B., White O., Beriman M., Hyman R.W.,
RA      Carlton J.M., Pail A., Nelson K.B., Bowman S., Paulsen I.T., James K.,
RA      Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA      Chan M.S., Nene V., Shallow S.D., Suh B., Peterson J., Angluoli S.,
RA      Parera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA      Martin D.M., Fairlaid A.H., Fraunholz M.J., Rhee D.S., Ralph S.A.,
RA      McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA      Venter J.C., Carucci D.T., Hoffman S.L., Newbold C., Davis R.W.,
RA      Fraser C.M., Barrell B.;
RT      "Genome sequence of the human malaria parasite Plasmodium
RT      falciparum."
RL      Nature 419:498-511 (2002).
DR      EMBL; AB014822; AAN37017.1; -.
DR      HSP; O931C2; IMW.
KW      Hypothetical protein.
SQ      SEQUENCE 3504 AA; 408308 MW; B8454D48D5B84F0 CRC64;

```

Query Match 6.6%; Score 264; DB 2; Length 3504;

Best Local Similarity 21.2%; Pred. No. 0.014; Matches 196; Conservative 131; Mismatches 346; Indels 252; Gaps 43;

```

Qy      11 KNLGKSGSLKKTDTGVHNNQEN--ESIKER-----SFTIDRNISTIRPFENKDL 62
Db      1432 KNSHNSSEDMITBEYGNKNSPOENIDMIIPKNGENKNSQNSDHNIMTEKNGKNS 1491
Qy      63 KKLKKRREVDPTSETGRMEEDYKTDK---GNLLADGDTLDEYTRKDLRIS 118
Db      1492 QQ-----HSDNIMTEEYKKNKNSQENTDNTBQYGNKNSKKKEEDIAS 1539
Qy      119 KIYGV-----LSPSK-GHPEILGKISVSKNAKVYV-----NNYSIRIKATKYP--- 165
Db      1540 --YEDKRIISHENQEHPTYESRINKGFHNVVDNIYSGNEKGISVKNISSEIIRD 1597
Qy      166 -HSKMTTDLVANIINDIYDGLAPADGRLFTVQNDOKKAEIKIRMEKIKETKS-YP 221
Db      1598 GKGKNIODDILLEN-----GKNKFBENIEBEDIKSDXTQ--SKSISHEABGHFTP 1647
Qy      222 YVS-----SYGVNIEIGEGDLKNNKPDNLTQMSGKIT-----SDSEKQGYLL 264

```

```

Db      1648 YESGKXKISDENVENISDINTNKOQ--EVESKQIFERNONINKHISSDNNKINKOK 1705
Qy      265 KNNIILRKGYALKVTTYNPKTD-KLENGCYKEDIKQKAPNLR-----ALSETTI 318
Db      1706 QNNILINESODKNI DVHN--KLDRIKNEVYTSDESLKKEKNGTRKNGKGINNEKI 1763
Qy      319 YADSRVY-----EDGSRQSVLMSALDGFNIIRYOVTFPKN----- 355
Db      1764 EEEKENVADDETLIGKEERTESDCLKIQKISNENKINILYTDNTNKKSYAAGGTHG 1823
Qy      356 -----DKGBAID-----KQNTLV----- 368
Db      1824 ENDETNGTNISNDGDKRVKTIQOYISKGNITLOENKBDIIPSVTINNSIGDKVXENLS 1883
Qy      369 -TDSSTLVFGHDKKRYTGDEKNV-----EAKKESQMLFDTK-----PVYLS- 412
Db      1884 PEDIKKAEVAHKNIQNTSEDELGTQKONERKEKSPGVEENQENDKIIGEVNLN 1943
Qy      413 MDKRYFNPSKNTIYVRNPEFYLRK-ISKGGFNNELR--VRESVVDN-----L 460
Db      1944 KNNNESNIGNSDTI---NQHLMGKNIHKKGVNSETMENTNSGTQNIISNGEPKNI 2000
Qy      461 IYGDLIHIDTRPFNITLAVKDGIDMDGKDYKANGPPKVTMDGNAVYLGTSYDLNAX 520
Db      2001 IRGD-DIKOMNENVKTLEDGTN-----NIKINYNDAKILNELIKNGQTSODAD 2052
Qy      521 AVGVHQLYDNVK-PEVNIIDPKGNTSIEYADKSVVFNINDRNNGFGEIOBHIYIN 579
Db      2053 DISTNSDKRDKQDQENNNENIHKINNVAE---KDKI--SNDKEDNIVPE-HKEDDIIS 2105
Qy      580 GKRYTFNDIKQIIDTFLNIKIVKDPARTYVKEFLINKDTEVELKPHRYTVTIQNG 639
Db      2106 DNKKKEFDNVLEIPKIGNI-----LDDKRTTBQVEEKSIGQDSMENNVSST--NDG 2157
Qy      640 KEMSTIVSEDPFLPVYKGLBKGYQPDGWEISGPEGKADGYVNLK----- 689
Db      2158 KQIH---IQEBDI-----KENIINNVDKHSKKNMIHIDPNKYVEKEIKKH 2204
Qy      690 ---DFTIKVVPKRIEKEEENKPT-----FQVSKKDDPQVNHQSLNESHREKDLQRE 741
Db      2205 EIAHDHDKKFKKELIQGDSNKNKSEPSNENILVDNAODK-----NISKLTNLDHDS 2256
Qy      742 HSQKSDSTQDVATVLDKNNISKS 766
Db      2257 KGTNDSVVE-----HNSDKT 2273

```

RESULT 8

```

ID      07RGP2      PRELIMINARY;      PRT;      1850 AA.
AC      07RGP2;
DT      01-MAR-2004 (TREMBLrel. 26, Created)
DT      01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DE      01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE      Hypothetical protein.
GN      Name=PY04304;
OS      Plasmodium yoelii yoelii.
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=73239;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=17XNL;
RA      Carlson J.M., Angluoli S.V., Suh B.B., Kooij T.W., Perea M.,
RA      Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA      Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell T.V.,
RA      Shallow S.D., van Aken S.E., Riedmiller S.B., Feldblum T.V.,
RA      Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,
RA      Florens L., Yates P.R. III, Raine J.D., Sinden R.B., Harris L.M.,
RA      Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA      van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA      Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,

```


QY 186 AFAGDMLPYKNDQ-----KKAETIKRMPK-----KETSRYPVSSYGNVIEL 232
 DB 773 -----TAKENDQNLVYKSKISEIKNHYNQDISIDNTYEEKKONYDOSKRYTITSI 824
 QY 233 GEGDLSKN-----KPDNLTKE-----SGKIVDSEKQYILDNIIIRKGYALKV 278
 DB 825 KEDSTKINIVANMDEFLSKYDKINFENCKEANDSHETQTELNTIKAEVSDK 884
 QY 279 TTY-----NPGKTMLEANGVYSK-----DIACIQRANPLRALST-----TYAD 321
 DB 885 SIYENKPNDSKSLINETNNSIEKRYONINTLKVDEIKYCESTYKBSIKNPHNKOTLIK 944
 QY 322 --SRNVEDGKSTOSVLSALDGFNIIRYQVFTPMQDNGRAIDQGL-----VTDSSLYL 376
 DB 945 KANQNKITIKETNISIEKSYDKP-----NLIKINELDKP-----KQASLNDYESNNELMQ 999
 QY 377 PGHDKREYTBDEKFNVAIKEDSMLPID-----TKPVLSMDK--NYNPKSKN--KIYVRN 430
 DB 1000 YFNNLKANLQKKN-----MLYHQDEKEKAVNNIIQKIEDINQILPIEIAIYT 1050
 QY 431 PERYLAGKISDKGFMNEL-----RVNNSVYDNYLIYGDIAIDTRDPRIKLVYDGD 483
 DB 1051 STYNISEELIENIGKINISLNTVLEKVTNTNPNFKIKELKLYNPSDFG-----KEGN 1105
 QY 484 IDWGMKDYKANGFPDKVTMDGNVYLYQYSDLNAAVGVHYQFLYDNYGPEVNIIDPK- 542
 DB 1106 I-----KYTEVANKIKODIKVNOQIDHINELEIKKSENYVDEKMAQINKLEKY 1157
 QY 543 -----GNTSIEVADGK--SVVFNINDRN-----NGFDGIQEQHYIYINGEYTSFNDIK 590
 DB 1158 ADTAISNDVNEGIEKQKQNIIVTKDKKKNTYBEINKLSISKLE-----KQGSLSKRYK 1212
 QY 591 QI-----IDTILN 598
 DB 1213 GIMLSYGONLGTFLBQIDBEKKSEHTITQAMEYMEDLNIIKKKSOEINENGIEMDIN 1272
 QY 599 IKIVNDPAMNTVKEBILNKDTGE--VSELKPRVYVTTIONGKMSSTIYSEDFILPVY 657
 DB 1273 KEMEVANISHDDKQVHTISKNNENISDIRNKSILKI-ID-----PSRSDINDI- 1322
 QY 658 KGELEKGY--QPDGHEISGFBGKQAGYINLSKDTPIPVFKLIEKKEBENKPTFDV 714
 DB 1323 KKELOKNSVBSQKNSBINQYLSK--IANYINILKINKIKQINKVEXYNEBIEK----- 1375
 QY 715 SKKQDPVNHQSOLNESHKEDLORBEHSQKSDST--KDYATATVD-----KONNISKST 767
 DB 1376 NKNQINDELSNEXLIKIKYKODISLCKECSKISSTLDDKQIDGCIQNIKVLKTHILSEET 1435
 QY 768 TNN 770
 DB 1436 NIN 1438
 RESULT 10
 QYREH9 PRELIMINARY; PRT; 1811 AA.
 AC QYREH9;
 DT 01-MAR-2004 (Tremblrel. 26, Created)
 DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE Hypothetical protein (fragment).
 OS Name=PY05085;
 OC Plasmodium yoeelli yoeelli.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XN;
 RA Carleton J.M., Anguiano S.V., Sub B.B., Kooij T.W., Perce M.,
 Silva J.C., Ermolaeva M.D., Allen J.B., Seleput J.D., Koo H.L.,
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,

RA Shallom S.J., van Aken S.B., Riedmiller S.B., Feldhym T.V.,
 RA Cho J.K., Quackenhush J., Sedegan M., Shoaibi A., Cummings L.M.,
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.B., Harris M.A.,
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
 RA van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Garucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 RT parasite Plasmodium yoeelli yoeelli";
 RL Nature 419:512-519(2002).
 CC CAUTION: The sequence shown here is derived from an
 CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL: AABL01001591; BAA17051.1; .
 KW Hypothetical protein.
 FT NON_TER
 SQ SEQUENCE 1811 AA; 210756 MW; E76A90F7392BCA0A CRC64;
 Query Match 6.4%; Score 258.5; DB 2; Length 1811;
 Best Local Similarity 21.7%; Pred. No. 0.011;
 Matches 179; Conservative 117; Mismatches 312; Indels 215; Gaps 40;
 QY 11 KNLGKRGSLKQDTTGVHRRHOENBS-----IKKSSFTIDRNISTIRDFENKD 61
 DB 555 KNINSGHGSNIHVRNRIHIDNNNDQNTVSNHILKQDETQNDIVDQNTATYGF 614
 QY 62 LKTLIKKKRREVDFTSETGRKHEBYKTDKKNIIAYDGTDLER-----ETKDE- 115
 DB 615 ILS--DKEDFFKLTFLPDLKMGSPFLPEKMLKDSKQWIDILQOREYINNNIELEKKEE 673
 QY 116 -----IKSKIYGVLPSEKDGHEIFILGKISVSKN-AKVYVGNVSKSIEIKATKTFPH 166
 DB 674 KKYLEKCIINNKIKELMCGRVKGLMNV--KLCNNEKKIINAFSLNIIKNNIHVTLIDYE 731
 QY 167 SKTYTP--DLYANINDIYDGLAFAGDKRLPYKNDQKAEIKIMPEKIKETTSYPIVSS 225
 DB 732 RSAMCYODIYA-----QLYVYGLNSK-----KIEMIKKINEK----- 764
 QY 226 YGANVIELGEGDLSKNKPDNLTKMESGK-----IYSDSEKQYILKNIIIRKGYAL- 276
 DB 765 --NNID-GVANSQENSDAKKQDSSKKKGYVSHLSIPKSKH--KON--KKKYSIN 815
 QY 277 ---KVTYNGKTDMLGNGVYSKEDIAKIQRANPLRALSETTIVASRVNVEDGKSTOS 333
 DB 816 NDTKQTT--EPSKQD-----NLAKAEHBRITANNNEDEGRKLEN 853
 QY 334 VMSALDGFNITIKYQVTFK--VNDK-----GEALIDQGLVTPSSKVLVFGKD--KEYT 385
 DB 854 KQKG-----NINRTVNTKQNLNEQYKQNSDETDRKSNIRTDKANNPNFGKNSYNEKT 908
 QY 386 GEDKFNVEAIKEDGSMFLIDTKPVNLSMDKVFNPFS-KSNKIYVRNPEF-----YLRGKIS 440
 DB 909 VEDK-----EIKETKNSNVYIYNSNSYENNSYLDNNSFTRKQNLKNSKSEIV 953
 QY 441 DKGFNWELEKRVNNSVDN--YLIYGDIAIDTRDPRIKLVYDGDIDWGMKQDYKANGFP 498
 DB 954 EKSNNSN--IKKKKVDMGSGSPFGSP-----FMFK-----SDYNNKKG-- 987
 QY 499 DXYTMDGNVYLYQYSDLNAAVGVHYQFLYDNYGPEVNIIDPKGNTSIEYADSKSVFN 558
 DB 988 SÖIESNNGSALVPPKSDDSNB--GSDYSFKYNNKEDKROMKCKNSKKSNNMAFKK--IPN 1043
 QY 559 INDKRNNGPFGEIÖ--EQHIYINGEYTSFNDIKQIIDTTLNIIKIVKQDPARTVYKEF 615
 DB 1044 LSRKKKKKKKGGDPSPLKSKEDKIKKQDNTDMSVENNEBK-----YNNSSNQ 1091
 QY 616 ILNKDTGEVSELKPRVYVTTIQG-----KEMSSITVSEEDTILPVYKGELEKGYPDG 669
 DB 1092 SSNEEDGHINNYKN-----NGMKKAKNKSNGNLNKSLSLEFKDFDEFEFGI-- 1142
 QY 670 WEISGFEKKQDAGVYNLSQDTPTIKPVFKLIEKKEBENKPTDVSKKQDPVNHQSOLN 729
 DB 1143 -----LKGEVDPHLIHDVDEKSIIFERFDHED-----FDIGIKNSNSCNSN--N 1187

Qy 730 ESHREKDIQREHRSQSDSTKDVAT-VLDKNNISKSTTNP 771
Db 1188 DSMSSD-----SNKSIROSQDTQITKTKSSSKSEKQCP 1224

RESULT 11

Q7RMS4 PRELIMINARY; PRT: 2661 AA.

AC Q7RMS4: 01-MAR-2004 (TREMELrel. 26, Created)
DT 01-MAR-2004 (TREMELrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Rhoctry protein (Fragment).
GN Name=py02104;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angitoli S.V., Suh B.B., Kooli T.W., Pertea M.,
RA Silva J.C., Brmolaeva M.D., Allen J.B., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.B., Riedmiller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabhi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.B., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdiva A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

CC EMBL; AAB01000575; EAA21526.1; -
DR InterPro; IPR011561; Pox_11_rel.
DR InterPro; IPR006499; ReticulocyteBP.
DR ProDom; PD014111; Pox_11_rel; 1.
DR TIGRFAMs; TIGR01612; 235KDa-fam; 1.
FT NON TER 2661 2661
SQ SEQUENCE 2661 AA; 312269 MW; 0BD02AD0B71253B CRC64;

Query Match 6.4%; Score 257.5; DB 2; Length 2661;
Best Local Similarity 20.0%; Pred. No. 0.018;
Matches 179; Conservative 151; Mismatches 294; Indels 271; Gaps 41;

Qy 22 KQDTTGEHHQNESEIK-EKSSFTIDRIS-----TRDPENK-- 60
Db 427 KKLLETTTKANKNEBYTLKEIRELFKYSDEBAEKGYBELKALKETITKIDYNNKE 486
Qy 61 -----DLKKLIKRRFVDDFTSRYGRMEBYDYKYD-----KGNIIAYD 101
Db 487 YIKKAIIDLKAIENNNKYIDELGKNTFPOIREYVKKOTIYSTIKSELISYKNIIVELY 546
Qy 102 DGTDLRETRKLDITK-----LYGULSPKRGHRELIGKISVSKR----- 144
Db 547 NEUSVVOENTTIDPIKKNKELETLSKSEIDYNNKIKNNKIBREVAHLAKNISTKNSL 606
Qy 145 -----AKVYVG--NNYKSIIEIKATKYDFHKTFTFDLYANINDYGLAFAGMRLFYVON 198
Db 607 ILLEIKYIFGEID-----DLNKITKEFNKQO--ELSKIND-----YTKEN 647
Qy 199 DQ-----KKAIRIKRPEKI-----KETYSRYPVSYSGVNIELGEGOLSK----- 239
Db 648 DQLSVYOSKIIIRKHYNDQINVDNTKGBEAKONYDOSKRYAKKIPIRGTSKFXINVT 707
Qy 240 -NKPNDNLTKR-----SGKIYDSEKQOYLKNNI-----LRGYALAKYTTNPG 284
Db 708 AMKNELLKJLKKYIDPGAYKAKANSSEKITELVKIRIRISKYLANNYERK--FNNS 764

Qy 285 KIDMLKGNVYSKE--DIKLIKQANPNLALSETTYYAASRVNEDGRSTQSLMSALD-G 341
Db 765 KKLITETDKSIEBEYKNINWTLKKVDYIYCVHT-----NELSNRNKQTITLKLDNRN 819
Qy 342 FNIRY-----OVFEK-----NMDCBAIDKGNLY-----TDSKGLVLF-----G 378
Db 820 INTVKENNSIDLTYNTKFNENILTDKTELDKPTTVALADHESNNNELKYPNNLKANLG 879
Qy 379 KDK-----RYTGEDKFNVEAIKDGSMLEFIDTPVNLSDKNYFNSKSKYVNRPEFY 434
Db 880 KKKKNMLYQOFGEKEKALIDIKKADINENYSKIEVALYASIIYNSIDEMERIEKSIES 939
Qy 435 LRGKISDKGFMWELRVNESVVDNTLYIGDLHDINRDNRNFKLNVQSGOIMKMDYYA 494
Db 940 LNTQYTER-----VKTNTNLNEIKKGLKQYDHDGKENIK----- 977
Qy 495 NGEPDVTMDGNVYLQTSYDLNKAAGVHYOFLDYNKPEVNIDPKGNTSIEVADGS 554
Db 978 --YDPEINKIKNEI-----KTVG-----QOIDQHINKLEIKKSGSKNDELKIKGT 1021
Qy 555 VVF--NINRRKNGPGEIOEHIYINGKEYSPN-DIKQIDKTLNKKYVOPADNTTV 612
Db 1022 DEFENVTDK-----ITYNKDPKEIEEKIKNI-VTKIDKKNIYI 1059
Qy 613 KEFLINKQGEVSEILKPHVT-----VTIQNGKMSSTIVSEDEPILFVYGELEKGYQ 666
Db 1060 K---INKLIDETSEIKNNTSLKKVKVDINVSYGOSIGKLFLEQID----- 1101
Qy 667 PDGWEISGFBGKKGAGYVNLNPKDTIFKVP--FKQIEKKEBKRPD-----VSKKKON 720
Db 1102 -----EKKKAGHNIK-SMEBYMDLNIKKKSGEIEBKIMNDIKYINKEMEA 1151
Qy 721 POWNSQLNESRKEDLQREHRSQSDSTKDVATVATVLDKN--ISSKSTTNP 772
Db 1152 LNISSDDKQYH--NISKGBEKISD-----IHNSIKTIQEFSTESNIN 1194

RESULT 12

Q7RPS4 PRELIMINARY; PRT: 1389 AA.

AC Q7RPS4: 01-MAR-2004 (TREMELrel. 26, Created)
DT 01-MAR-2004 (TREMELrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=py01465;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angitoli S.V., Suh B.B., Kooli T.W., Pertea M.,
RA Silva J.C., Brmolaeva M.D., Allen J.B., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.B., Riedmiller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabhi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.B., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdiva A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.

DR EMBL; AAB01000391; EAA20892.1; -
DR InterPro; IPR011591; Botulinum.
DR ProDom; PD001963; Botulinum; 1.

KV Hypothetical protein.
SQ SEQUENCE 1369 AA; 164816 MW; 21851CA630F5D9F CRC64;

Query Match 6.4%; Score 257; DB 2; Length 1369;

Best Local Similarity 20.4%; Pred. No. 0.0089;
Matches 193; Conservative 139; Mismatches 288; Indels 326; Gaps 48;

```

QY 17 KEGSLKQDTGVVHHHNOE-----NESSIKKSSFTIDRNSTIND 56
DB 354 KKEAFQKKRNSSEHLEKNIPIYKESILKASKTDTVSYPFKSEIKKKSSNNYNNKGVIS 413
QY 57 FENNOLKLLIKKRFVDDFTSETGRMEEDYKYDDKGNIIAYDDGTLEYTEKLEDEI 116
DB 414 EIKKNGQIKKNCNKHVFSHSDT-----EKVEY-----SNIS-----DLVNGKNDX 457
QY 117 KSKYGVLSPSKQGFELIGKSNVSKAKAYTGG-NYSISIKAT-----KY 163
DB 458 KLMNYGL---SKEDY---SNDKNNQDTSEYVYDNEBNDKTDIDIOVNDKKNVLDISPLYTHY 510
QY 164 DFHSK-TWTPDLVANIINDVGLAFAGMRLFYKNDOKKAEIKIR---MPEKIKETSE 219
DB 511 KYRNNIITINKIKNIKI-----EINAMITTVOLKIPGRIENLVLPKH----- 556
QY 220 YPVYSSYGNVI-----ELKGGDLSSKNKPDNLTKMESGKIYSDSEK-----QOYL 263
DB 557 -----YENDLIKQIKTISEENENDIKTNKE---TKNATSYYQSDKNSIQPSNDNV 607
QY 264 LKDNIIILKGYALK-----TTNPGKIDMLBNGVYSKED- 299
DB 608 IKKNNIDKTSIKIYVILPANNINHEIINRLLINSTYB-----MYEKENTL 657
QY 300 -----IAKIOKANPMLFALSETTYADSRNVEDSGSTQVL-MSALDGFN---IIRYQ--V 349
DB 658 LPTDTYILQ---SKSISIKKEMMYISIKKCVNGIYLPILANNVLDNFIRKNNMISQNNN 715
QY 350 FTFK-----MADKBAIDKO-GNLYTDSKLVLFKQDKKYEYGDKNVBAIKEDGML 402
DB 716 STFKPYVSLYNDK---IDREIGMLI-VTNKLI-----NKIKIKKLPKGGYIT 759
QY 403 -----FIDPKPYNLSMDKN-----YENPSKNTKY 427
DB 760 INDICFRFPDKNNVLCVYNNRKSNSGNGIQQMLAIDHIGVETANYIFMAGSSNFTI 819
QY 428 VRNP-----EFTLRKGISDKSGFNMELRVNBSVDNVLTYGDLHIDN 469
DB 820 VKSSLDLPFPHIFLLIBQVYVYDFIEKNEMLLAKFIMHPLAITBAICINLCFL----- 872
QY 470 TRDNRIKLANVQDGMKGMQDYKANGPDDKVTMDGNVYLOTGYSDLNAKAVGHYQFL 529
DB 873 ---FNINAQI---DSLFLY--KMLKAVQIKENDQNNPFLTYLSYB--RVQNNNGHCEYF 923
QY 530 YDNVSP-----EVNIDPKGNTSIRYAD--GKSVFENIDKNNNGPDGEIOBOHIYI 578
DB 924 LNVSKPLFRPOSSEHLA-----NISHADVEGHTNLTGNTSDERIKCEBIIKKCKEIE 976
QY 579 NGKETSFNIDIKQIIDLTKLNIKVVDPAFNTVVEPLANKDTGVSBLKPHRVTVIQN 638
DB 977 NG-----IK-LIKKTNDNDPFIYDAPKQSI--FLVN-----AIKNDALIKNIEN 1018
QY 639 GKEMSSITVSEEDF-----ILPVYKGEHLK-----GYOF-----DGM 670
DB 1019 NYDKKKQIFTHYSTYKQPODYIINILNRVNOQLQKQKEDYSRRREVETLQSNIRPQ 1078
QY 671 EISGFEKQKAG-VYINLSKDTFIKPFKLEKKEB-----ENKPTFVSKCKQNPVN 724
DB 1079 DVKSYFGKNDNNITPSGGLNQ-----FDKCKEKKRBGLIKSSNNKMGQNDNNTYDNYQS 1131
QY 725 HSQANESHRKEDLQREHSQKSDSTKDYATATVLDKNNISSKSTNN 770
DB 1132 KQTESDKKPFQNIYTDENIQMLN-----NNSNTNQVNN 1165

```

RESULT 13

```

O813P4
ID O813P4 PRELIMINARY; PRT; 1777 AA.
AC O813P4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE Hypothetical protein PF01095w.
GN Name=PF01095w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Alkin R., Baker S., Barton A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphrey S., Jagsels K., James K.D., Johnson D., Kethnorou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultana J.B., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL929353; CAD51583.1;
DR InterPro; IPR011591; Botulinum.
DR Prodom; PD001963; Botulinum; 1.
KW Hypothetical protein.
SQ SEQUENCE 1777 AA; 213320 MW; 244467CFPI90C522 CRC64;

Query Match 6.3%; Score 255; DB 2; Length 1777;
Best Local Similarity 20.7%; Pred. No. 0.015;
Matches 186; Conservative 152; Mismatches 324; Indels 236; Gaps 44;

QY 4 EIAESKPNLNGKSGSLKQDTGVVHHHNOE-----ESIKKSSFTIDRNSTIDPF 57
DB 977 ELLDDKKKKL--DEENELDDKK--KGLDEENELDDKKKGLDEENELDDKKKKL--DE 1030
QY 58 ENK--DLKLL-----KKKFRVDDFTSETGRMEEDYKYDDKGNIIAYDDGT 104
DB 1031 ENEELDDKKKGLDEENELDDKKKGLDEENELDEERKKKMDNDNILLDEKKGEIVNDHT 1090
QY 105 DIEVETKLDIEIKSI---YGVLSPS-IDGHE-----EILKISVNSKNK----- 146
DB 1091 FQTEHNLINKEKTKIQIDYNIIEALIKEMISKILLSEETKIKINDENDIDIKRQSOIEN 1150
QY 147 VYGGNNYSIEIKATKYDHFSTWTFEDLYANINDVLDGLAFAGMRLFYKNDOKKAEIK 206
DB 1151 IYRRNSMMIDNNISYSKSNVTKTFNPDNNTE----- 1184
QY 207 IEMPEKIKETSEYVSSYGNVIELSGGDLSSKNKPDNLTKMESGKIYSDSEKQOYLTD 266
DB 1185 ---EDKIQKNTKITYDDISM-----LTNNKSSGISYNSKDIINNOEIIYSNG 1234
QY 267 NI-----ILKGYALKVTTYNPDKTDLBNGVYSK-----EDIKIQK-ANPN 309
DB 1235 NTFNNDIEKDLTNVAKNNNFNNISYKQCSNIVYINNNKFLMDTSESIKIQNDINER 1294
QY 310 LRALS--FTTYADSRNVEDSGSTQVLN-SALDGFNIIRYOVTFKNDKGEALDKQCN 366
DB 1295 VTLKQREEIFYEEKGNFEKNEHLAKENVLVNQNIIIKOR--ENEAKKEEELKQKGE 1352

```

```

QY 367 LVTSDSKLYF-----GKODKEY-----TGEDKENYBAIKEDGSMLEPIOTKPYNLGM 413
DB 1333 ILLSEKENTINYSLSNCKINKKLDKLETKLEADVTGKGDINNSYMKSPKSHL--SDTN--NIOP 1409
QY 414 DKQYFNPSKSNKIYYRNPEFYLRGKISDGKGFNMBLAV-----NESVDNYLLY-G 463
DB 1410 FENDIN-----NDKLYEDNISSTY--GNNLSHNNKMYDNNSLCDNNMSHDNMADNNI CHN 1465
QY 464 DLHINTEDFNITKANV-----KQGDINDGKQYFANGFPDXYVTMDQGVY----- 509
DB 1466 NISYNNPNSYNTHMKLSBELBNENKKNKONIDY-----YDKNFVDSNNININHNFNISKI 1521
QY 510 ---LQTSYDLNAKAVGHYQFLYDNYKPEVNI DPKGNSTIEYADSKSVFENIDRRNG 566
DB 1522 SMDISISGKNDMB-----LHDKLV-----NNSLKEIETMYKLSRDESEINT 1562
QY 567 PDGEIOBQ-----HIYNGK-----EY-TSPNDIKOIIDKTIN 598
DB 1563 LKLKLEKQRTBEKMKONNINQANDISDMNNIYINGNISGVYVOTSMHVLBELPERKFLN 1622
QY 599 IKIVKQDPARNTTVEFLINDTG---EYSELKPHRYVTTIQNGKMSSTIYSBEDFT-- 653
DB 1623 LS-EVED---NEKMSIISNELVSLYKELNNIK-----BEYNINVLKQEFIGN 1667
QY 654 -LPVYKGELEKGYQDGEWEISGPEGKQAGVYINLSKOTPIKPYFKKIEBKKEBEKPTF 712
DB 1668 LLINFLANDLNNRYL---KANYEKEGVHKSQIILISREDFIKELQNTLMBKELKE----- 1719
QY 713 DVSKKQDPQVNHSQLMESHK---KEDLOREBHSOKSDSTQVATVATLQNNISKSST 767
DB 1720 -ISYKQMLKKN--QINDTYKLNKRKSLSTVELKQ--DIKFLNEBVLKCKEAVTLST 1772

RESULT 14
QYVT6 PRELIMINARY; PRT; 1127 AA.
QYVT6
AC QYVT6;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein MSV156.
GN Name=MSV156;
OS Melanoplus sanguinipes entomopoxvirus (MeSPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.
OC NCB1_Taxid=83191;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=99102612; PubMed=9847359;
RA Afonso C.L., Tulman R.R., Lu Z., Oma B., Kutish G.F., Rock D.L.;
RT "The genome of Melanoplus sanguinipes entomopoxvirus."
RL J. Virol. 73:533-552(1999).
DR EMBL; AF063866; AAC97677.1; -.
DR F063866; AAC97677.1; -.
DR Hypothetical protein.
SQ SEQUENCE 1127 AA; 134265 MW; F18SDALD5A3FE7D1 CRC64;

Query Match 6 3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0098;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;

```

```

QY 162 KYDFHSKMTFPLXANINDIVDGLAPAGMELFVK--DNDQCKA-BIKTIMPEKIKET-- 216
DB 368 LFDNDIQKLANDITTEONNKITD--FPNNSTRIFKEKLDTEYKLDIDIKNNNLQCKLESYK 425
QY 217 ---KSEY---PVSSGVAVTELGEGDLKKNRPDNLTKMSGKIDGSEKQOYLAKONII 269
DB 426 KIDQETVEYKNNKINNEYNDIIEKNNNLQCKLEBNCKIDQETVEYKNNKINKEV--NDIIE 483
QY 270 LKRGYALKVTTNPG-----KTDMLBGNVYSKEDIKIQKANPNIPALSETTYAD 321
DB 484 LKNNNLQCKLEBNKININDLTKLNDIESENTBELFNLANISDFQKREIRIAKLN--TEYEQ 541
QY 322 SRR--VEDRSTQSVL-----MSALDGNRIIRYQVFTFRNDKGAIDQXGLVYDSSK 373
DB 542 LKNDLNTNKTRELKLSDNKLSLE-----QLYDSKRL--DGDILYNSLTKCN- 592
QY 374 LVLFGKQDYKRYGDFNVEAIKED---GSMLEPIDKPYNLSDMKYFN---PSKSNKIY 427
DB 593 ---DKIDYFNSIEKFDIYVYENKFIQNDLSINKIINNDFKREYINSKIDSKNELS 648
QY 428 VARNPEFY-----LRKISDGKGFNMBLAVNESVVDNYLLYGDHLINTRD 472
DB 649 TWFDDIFNAKQJASITNNIENISNKIXD-----INFEIISNEDSKEL-LDEIRK 698
QY 473 FNILKLVKQDIDMDGKQYKANGFPDXYVTMDQGVY-LQTSYDLNAKA-----V 522
DB 699 YKQQPD-KIKDANNEVKSFE--NTLQXQIDISKSNBELTNADVIINTQANDLDDKLVNY 756
QY 523 GVHYQFLYDNYVP-----EVNIDPKGNSTIEYAD-GKSVVFNINDRRNGPDGEIOEHI 576
DB 757 GSBFKNLYNNASDLDDITQKNNDEKQKQNTLEKKNQSIENIDVNN----- 805
QY 577 YINGKYSFPNDIKOIIDKTINIKIVKQDPARNTTVEFLINDTGVEB--LKPHRYT 633
DB 806 FI--KELIKFNTE--TNKSLN-ELLTND---DINDKIKLYKELNKISTNNILKTYK- 855
QY 634 VTIQNGKMSSTIYSBEDFTLPVYKGELEKGYQDGEWEISGPEGKQAGVYINLSKOTF- 692
DB 856 NEDIDVNEGLSIVENLOFINFLSIEFQO-----STSH-----INFLANTLA 900
QY 693 -IKPVFKIEBKKEBEKNTFPVSKKQDPQVNHSQLMESHKEDLOREBHSOKSDSTQ 751
DB 901 GINDVLANKLNLKIMADTTRGQVNIIR-----DEIKQISSENKSKQFNEKNE--KD 950
QY 752 VTAIV-----LDKNNISKSSTTN 770
DB 951 LKGLISFNDKLNKRYNSAGYTEYN 974

RESULT 15
QYVT6 PRELIMINARY; PRT; 2227 AA.
QYVT6
AC QYVT6;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFName=PF11_0354;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCB1_Taxid=36329;
RX [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung B., White O., Bertzman M., Hyman R.W.,
RA Carlson J.M., Pahn A., Nelson K.B., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyte S.,
RA Chan M.S., Nene V., Shallow S.J., Sub B., Peterson J., Angioli S.,
RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairhead A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2005, 19:11:30 ; Search time 44 Seconds
(without alignments)
1690.354 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGIABSKFKNLGNKGKGS.....ATVLDKNNISKSTNNPK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2140	2 P95074	serine proteinase,
2	3709.5	92.1	2144	2 A97942	metalloproteinase,
3	253.5	6.3	1127	2 T28317	ORF MSV156 hypothe
4	250	6.2	1650	2 T18444	hypothetical prote
5	235.5	5.8	3724	2 T18427	hypothetical prote
6	231.5	5.8	2269	2 T28677	rhodopsin protein -
7	230	5.7	2401	2 T18476	hypothetical prote
8	228	5.7	4550	2 T18440	hypothetical prote
9	221.5	5.5	1711	2 T18429	hypothetical prote
10	219	5.4	1640	2 A24594	probable major sur
11	218	5.4	1631	1 SAZOK1	major serine prote
12	218	5.4	1639	2 S05603	lipoprotein (impor
13	217.5	5.4	1546	2 G90603	reticulocyte-bind
14	215.5	5.4	1252	2 B42771	DNA-directed RNA p
15	214	5.3	2339	2 A45597	membrane nucleas
16	212.5	5.3	1125	2 E90598	hypothetical prote
17	212.5	5.3	2166	2 G70163	serine/threonine-s
18	212.5	5.3	2485	1 H71621	protein with 5'-3'
19	212	5.3	1188	2 A71621	microbial collagen
20	211.5	5.3	1104	1 A36866	RBSA-H3 antigen pr
21	209	5.2	1558	2 B1603	hypothetical prote
22	207.5	5.2	4688	2 P82885	hypothetical prote
23	207	5.1	1141	2 B89824	chromosome segrega
24	207	5.1	1191	2 B97116	hypothetical prote
25	206.5	5.1	1979	2 C71622	probable cell surf
26	206.5	5.1	1622	2 AB1717	mature-parasite-in
27	206	5.1	1526	2 A45605	RAD2 endonuclease
28	204.5	5.1	1516	2 E71619	hypothetical prote
29	204.5	5.1	2500	2 G71609	hypothetical prote

30	204	5.1	1245	2 D71613	GAF domain protein
31	203.5	5.1	3394	2 T18501	hypothetical prote
32	203	5.0	1166	2 T28680	fibrinogen-binding
33	203	5.0	4981	2 T18489	hypothetical prote
34	202.5	5.0	763	2 G97026	superfamily 1 DNA
35	202.5	5.0	1939	2 T18372	repeat organellar
36	201	5.0	1712	2 C71618	hypothetical prote
37	199	4.9	888	2 E71608	ATP-dept. acyl-CoA
38	199	4.9	980	2 E71606	hypothetical prote
39	197	4.9	1386	2 AC1533	surface protein (l
40	196.5	4.9	2523	2 T18477	hypothetical prote
41	196	4.9	1714	2 E71609	Ser/Thr protein K1
42	195.5	4.9	2380	2 E71604	hypothetical prote
43	195	4.8	1008	2 T18508	hypothetical prote
44	195	4.8	1302	1 JG6009	surface-located me
45	195	4.8	1365	2 T30822	Impl protein - Myc

ALIGNMENTS

RESULT 1

P95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIG
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #ext_change 09-Jul-2004
C/Accession: P95074
R/Retain: H.; Nelson, K.E.; Paulsen, I.T.; Bisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holczapf
nson, T.; Hickey, E.K.; Holt, I.B.
Science 293, 498-506, 2001
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venier, J.C.; Dougherty, B.A.; Morris
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:21357209; PMID:11463916
A/Accession: P95074
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-2140 <KUN>
A/Cross-references: UNIPROT:Q97RY6; GB:AE005672; PIDN:AAK74791.1; PID:G14972117; GSPD
A/Experimental source: strain TIGR4
C/Genetics:
A/Genes: SP0641

Query Match 100.0%; Score 4026; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.9e-164;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KLGIABSKFKNLGNKGKGSIKKDTTGVHHHONESTIKKSSPTIDNISTIRDPENK 60
DB	1334	KLGIABSKFKNLGNKGKBSLKKDYGVHHHONESTIKKSSPTIDNISTIRDPENK 1393
QY	61	DLKKLKKKFRVDDPTSETGRMBEYDYKDDKNTIAYDGTDLLEYETKDEIKSK 120
DB	1394	DLKKLKKKFRVDDPTSETGRMBEYDYKDDKNTIAYDGTDLLEYETKDEIKSK 1453
QY	121	YGVLSKDGHPFELIKISNVSKNAKVVYNNYKSIETKATYDFHSKTTFDLYANIND 180
DB	1454	YGVLSKDGHPFELIKISNVSKNAKVVYNNYKSIETKATYDFHSKTTFDLYANIND 1513
QY	181	YVDGLAPADMLPYVDNQKAEIKIRPEKIKETKSYPPVSSYGVNTEIGDGLSKN 240
DB	1514	YVDGLAPADMLPYVDNQKAEIKIRPEKIKETKSYPPVSSYGVNTEIGDGLSKN 1573
QY	241	KPDNLTQMSGKIYSDSEKQYLLKONIILKGYALKVTTYNPKTMDLGNVSKEDI 300
DB	1574	KPDNLTQMSGKIYSDSEKQYLLKONIILKGYALKVTTYNPKTMDLGNVSKEDI 1633
QY	301	AKIQANPVLRALESTTIYADSRNVEDGSTSVLSALDGNVILRYQVTFPMNDKGA 360
DB	1634	AKIQANPVLRALESTTIYADSRNVEDGSTSVLSALDGNVILRYQVTFPMNDKGA 1693
QY	361	IKDQGLVYDSSGLVFGKDDKEYTGEDFENYBAIKEDGSMLPITDKPVNLSDKKYFNP 420

Db	1694	LDXGNI,LTDSKL,VL,PKDQKXBXTGBDK,FNVBA,I,KEDGSHL,PLDTRKVN,LSMDXVFN,P	1753
Qy	421	SKSNKI,YVRNBE,FLRGK,LDKGGFNMEL,RVNESVVDNY,LYGDLH,INTDFN,IKLVNK	480
Db	1754	SKSNKI,YVRNBE,FLRGK,LDKGGFNMEL,RVNESVVDNY,LYGDLH,INTDFN,IKLVNK	1813
Qy	481	DGDI,MDMKQ,KYKANG,FPDKY,LTDMDSNY,LT,OTGVS,DLNAK,VGHYQ,LY,NVNP,BEVND	540
Db	1814	DGDI,MDMKQ,KYKANG,FPDKY,LTDMDSNY,LT,OTGVS,DLNAK,VGHYQ,LY,NVNP,BEVND	1873
Qy	541	PKGNTS,LEADGKSV,FNIN,DRKNG,FDGB,I,QBOHI,YINGKBYTS,FNDI,KO,II,DKTLN,IK	600
Db	1874	PKGNTS,LEADGKSV,FNIN,DRKNG,FDGB,I,QBOHI,YINGKBYTS,FNDI,KO,II,DKTLN,IK	1933
Qy	601	I,IVVD,PARNTT,VXKE,FLIN,KOT,GEVSEL,KPHE,VY,LT,I,QNGKMS,STI,VS,BED,PI,LPVYKGB	660
Db	1934	I,IVVD,PARNTT,VXKE,FLIN,KOT,GEVSEL,KPHE,VY,LT,I,QNGKMS,STI,VS,BED,PI,LPVYKGB	1993
Qy	661	LEKGYQ,PGM,ITS,GFEG,KKDA,GY,IN,LS,KOT,PI,IP,VFK,IE,KKGEER,KP,FFDV,SKKDN	720
Db	1994	LEKGYQ,PGM,ITS,GFEG,KKDA,GY,IN,LS,KOT,PI,IP,VFK,IE,KKGEER,KP,FFDV,SKKDN	2053
Qy	721	PQVNSH,OLNE,HR,KED,LORE,BSH,OKSD,ST,KOY,TA,VL,DXNNI,SS,STNN,NPK	773
Db	2054	PQVNSH,OLNE,HR,KED,LORE,BSH,OKSD,ST,KOY,TA,VL,DXNNI,SS,STNN,NPK	2106

RESULT 2

Metalloprotease (NC 3.4.21.-) A (imported) - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
 C:Accession: A97942
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E
 y, R.; LeBlanc, D.D.; Lee, L.N.; Lefkowitz, E.U.; Lu, J.; Matsushima, P.; McAhren, S.; M
 y, P.; Sun, P.M.; Winkler, M.B.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Tashimas, S.R.
 A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11543234
 A:Accession: A97942
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2144 <CUR>
 A:Cross-references: UNIPROT:Q8DQ7; GB:AE007317; PDB:1G15458138; GSPDB:G
 C:Genetics:
 A:Gene: pTA
 A:Keywords: hydrolase; serine proteinase

	Query Match	92.1%;	Score 3709.5;	DB 2;	Length 2144;	
	Best Local Similarity	92.4%;	Pred. No. 6.4e-151;			
	Match	Local 714;	Conservative 26;	Mismatches 35;	Indels 1;	Gaps 1;
Qy	1	KLGRIASPKFKNLGNGKSGSLKKDTTGVHHHGHENBSIKKKSFTPIRDNISITIDFENK	60			
Dh	1339	KLGRISBDKLNKAKEPEENT-NNNQAKDESDPKSKSSVEGSAHLINNTIISTIREFFNK	1397			
Qy	61	DLKKLIKKKPRBYDDPTSETGKGMEYDYKYKDDGNIIAYDGDGLBATEKLDLDEIKSI	120			
Dh	1398	DLKKLIKKKPRBYNDPTSETGKIIEEYDKYDDKNIIAYDDGSLAQYTEKTFEDIKSKI	1457			
Qy	121	YGVLSPSKDGHEFIILGKINSVSNKAAKYGGANNYSIIEIKATYPDHSKMTFDLYANIND	180			
Dh	1458	YGVLSPSKDGHEFIILGKINSVSNKAAKYGGNSYKSIEIKATYPDHSKMTFDLYANIND	1517			
Qy	181	IYDGLAFADMDRLFYVKONDQKAEIKIRMPKIKIKETSEYPVVSYGAVYIELGEGLSKNI	240			
Dh	1518	IYDGLAFADMDRLFYVDNDNQIKAEIKIRMPKINKETTAEEYPVVSYGAVYIELGEGLSKNI	1577			
Qy	241	KPDNLTKMMSGKIVSDESKOQYLKDNIIILKGYALKTYTNPGKTMLBEGGYISKEDI	300			
Dh	1578	KPDNLTKMMSGKIVSDESKOQYLKDNIIILKGYALKTYTNPGKTMLBEGGYISKEDI	1637			

Qy	301	AKI OKRANPNALASETTIYADSRVAVGEGRSTQSLMALJOFENIIRVOVFFKANDGGEA	360
Db	1638	AKIOKAPNNAKYVASETTIYADSRVAVGEGRSTQSLMALJOFENIIRVOVFFKANDGGEA	1697
Qy	361	IDKOGNLTWDSKLVLFPGKODKEEYTGSDKTVNVEAIKEDGSMFLPDTYKPVNISMDKATPNP	420
Db	1698	IDKOGNLTWDSKLVLFPGKODKEEYTGSDKSVNVEAIKEDGSMFLPDTYKPVNISMDKATPNP	1757
Qy	421	SKSKKIIVVRNPEFLRKGISIDKGGFNMELRNBSVUNDNYLTYGDLHIDNTRDPMIKLVNK	480
Db	1758	SKSKKIIVVRNPEFLRKGISIDKGGFNMELRNBSVUNDNYLTYGDLHIDNTRDPMIKLVNK	1817
Qy	481	DGDIIMDGKMDYKANGFPDKVTDMDGWNVYLTQYGSJDLNAKAVGHYQFLYDNVPEVNIID	540
Db	1818	DGDIIMDGKMDYKANGFPDKVTDMDGWNVYLTQYGSJDLNAKAVGHYQFLYDNVPEVNIID	1877
Qy	541	PKGNTSIBYADGKSVFNINDKRNNGFDGBIIOBHIIYINGKEYTSPNDIKOIIDTITANIK	600
Db	1878	PKGNTSIBYADGKSVFNINDKRNNGFDGBIIOBHIIYINGKEYTSPNDIKOITDKITLINIK	1937
Qy	601	IYVMDPARNITVKEFLINAKDGEVSELKPRHVYTIIONGKEMSTIYSEEDFILPVYKGE	660
Db	1938	IYVMDPARNITVKEFLINAKDGEVSELKPRHVYTIIONGKEMSTIYSEEDFILPVYKGE	1997
Qy	661	LEKGYQPDGMEISGFEKGKDAGYVINLSKDTPIKPVFKIIEKKKEBENKPTFVDSKKDN	720
Db	1998	LEKGYQPDGMEISGFEKGKDAGYVINLSKDTPIKPVFKIIEKKKEBENKPTFVDSKKDN	2057
Qy	721	PQVNHSQLNBHREKEDIQREHSGKSNSTYDQVATYVLDKNNISSEKSTYNNPNK	773
Db	2058	PQVNHSQLNBHREKEDIQREHSGKSNSTYDQVATYVLDKNNISSEKSTYNNPNK	2110

RESULT 3

ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C:Species: Melanoplus sanguinipes entomopoxvirus
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C:Accession: J28317
R:Afonso, C.L.; Tilmann, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A:Reference number: Z20484; PMID:93102612; PMID:9847359
A:Accession: J28317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1127 <AFC>
A:Cross-references: UNIPROT:Q9YVT6; EMBL:AF063866; NID:94049647; PIDN:AAC97677.1; PID:
Genetics:
;Note: MSV156

[illegible]


```

Db 426 KIDEQTEYYKINKINRYNDIIELEKNNLQKLEENKIDQTEYYKINKINKEY--NDIIE 483
Qy 270 LRKVALKATTVYVPG-----KTMLENGVSKEDIQAKQKRNPMRLASETTIYAD 321
Db 484 LKNNNLQKLEENKININDKLTTLKNDISNTELEFNKLINISDKOKSRIAKLN--TEYBQ 541
Qy 322 SRN--VEDGRSTOSVL-----MSALDGFNIIIRYOVFTPKMDKGEALIDKQNLVYDSK 373
Db 542 LRKQLEENKINTELMKLSDNKLSLE-----QLYOSKQVIL--DGIDKIYNSLKERN- 592
Qy 374 LVLFQKQDKERTYGEDKFNVEAKED---GSMLEFDTKRYNLSMDKNTYR--PSKSKTY 427
Db 593 ----DKIDEYFSNTEKFDIYNYVTEENKFIQNLDSIINKIINDQFKKEYINSKIDSKEBS 648
Qy 428 VRNPEFY-----LRGKISDKGFGFMELRVNKGVSQVNYLYIGDLIHIDNTRD 472
Db 649 TMLPDDIFPAKQOIASITNNIENISNKTID-----LNEFTISNESSKEL--IDERK 698
Qy 473 FNILKLVKQGDIMDMGMDYKANGPPDKVTTMDGNV--YLQTYSDLMAR-----V 522
Db 699 YKQQPD--KIKQAMNTEVKSFR--NLLQKQIDISIKSNINELTNAYDIINTKANDLDKLANV 756
Qy 523 GVHQLYLDNKP-----EWNIDPKGNTSIEYAD--GKSVFPIINKKNNNGPDGEIQOCHI 576
Db 757 GSEBKALYNNASDLLDTIYQKNNDEKVKQINLEYLKKRQOSIEIINDIVNN----- 805
Qy 577 YINGEYTSFNDIKQIIDKTINIKIVKDPFANTVKEFPIINKDGVSEB--LKPFRVT 633
Db 806 PI--KELIKPNTS--TKSLN--ELTND---DINDKIPLKYELKANKISTNNLKIYK-- 855
Qy 634 VTIONGKMSSTVSEEDPILPVYKGELEKGYQPDGMEISGFBGKQDAGVYINLSKTPF- 692
Db 856 NEINQVNEKLSIVIENTQFINSPLSIEFNQD-----SITSH-----INFLANTLA 900
Qy 693 -IKDPFKKIEKKEBEKNEPTFQVSKKKONPVNHSQLESHEKEDLQREHSQKDSKTD 751
Db 901 GINDVNLKLNKLNIMADVTRRGDTNIR-----DEIKNQISSENIKQKNEKE--KD 950
Qy 752 VTATV-----LDKNNISSKSTNN 770
Db 951 LKQDISPDKLNKNTNISAGYTEYN 974

RESULT 4
118444
hypothetical protein C0385c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18444
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18444
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1650 <LAW>
A:Cross-references: UNIPROT:O77328; EMBL:Z98547; NID:e1325376; PID:e1427940; PIDN:CA8111
C:Genetics:
A:Map position: 3
A:Introns: 1597/3; 1625/3
A>Note: C0385c

Query Match 6.2%; Score 250; DB 2; Length 1650;
Best Local Similarity 20.8%; Pred. No. 0.004;
Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;

```

```

Qy 130 GHPEILGKS-----NVSNAKVVYGNNTYSI--EIKATKVDPSKTMTEFDLYANINDIV 182
Db 502 DEMELYNQHTDPAINENLANKIYF--DYBQDPAKAKKKKLDHLYTQOKKRNINDIL 560
Qy 183 DGLAPAGMRLFVKQNDQKAEIKIRMPBKIR-----ETKSEYPVYSSGVNIELEG-- 235
Db 561 K-----DHLNADKETEKQNEIEIEEKQNKIRIEEEKQNKIRIEEEKK 614
Qy 236 ---DLSKRPDLTNGESGKIYSQSEKQYLLKNII-----LR 271
Db 615 KKIIEEKKKKKI--EKEEKKKKIDDEKQNTYANDKISHIDVNCNIIKIDALLHIEEK 673
Qy 272 KGYALVTVYVPGKT-----MLBNGVYSKEDIK-----IQKAPNL-- 310
Db 674 KTGHEIMLYLKEIKIEYQKMLNDENSIMLEHEKKNTHTQVNNNLDCTQMLQKEXKILYN 733
Qy 311 RALSETTYADSRVNDGRSTQSVLMSALDGFNIIIRYOVFTPKMDKGEALIDKQNLVTD 370
Db 734 NDKKKTFLSKSKNI-----TSNVLSSKI PG-----TLSTCKLMAVTKTIKCD---VTD 779
Qy 371 SKKVLPG-----KDDKE--YTGEDKE--NYEA-----IKEDGSMLEFIDTK 407
Db 780 NEKKKYVDHDKQNIKKRKEPINITKGRNAYVEIGSBVCNKNVYKAGDQKQKQVENEK 839
Qy 408 P-----VNLMDKQNF-----NPSKSKNIYYRNPFR 433
Db 840 QGDDNNMVAENKQGDENNVIYVNIERSSESPFTHRRKNTSTDTCAKNEQKYPHL 899
Qy 434 YLRGKISDKGFMELR--VNESVDNY-----LIGYLAHDN----- 469
Db 900 Y--SNQBEDKDKSIFLKNINENIKKQYDKERISTLEKKVFVKQKNVITNNDDEKHTSK 958
Qy 470 -TRDENI-----KLN----- 478
Db 959 INDDPNTIYIDKQKGLANNPPDLAKKIKNETKILEKDSHNSKIQNNLMKKKTNPPTNKK 1018
Qy 479 -VKQGDIMDMGMDYKANGPPDKVTTMDGNVYL-----QTG 513
Db 1019 GISSTSISSSTQNKQDCGIEKKNKLNLSLKTYCIRKRNQVDSIKLNDKADLYKDKKTS 1078
Qy 514 YSDLMARVGVHQLYLDNVPREVNIDPKGNTSIEYADGKSVFN--INDKRNNGPFG 569
Db 1079 FNDINRAKGNVFK-----KRDV--PRKNMMDVTNKGK--VFNPVTLANNRYNN-- 1124
Qy 570 EIOBQIYINGEYTSFNDIKQIIDKTINIKIVKDPFANTVKEFPIINKDGVSEBSELP 629
Db 1125 -----YIRSNK--NNVK-----NGKNV-----GIKKIYLLKE--KQKSLHP 1156
Qy 630 HRVTVTQNGKMSSTVSEEDPILPVYKGELEKGYQPDGMEISGFBGKQDAGVYINLSK 689
Db 1157 EGVEADKFLANSVYNKYLEKQFPDIIINEEKEK-----YKNNK--MKYKI---K 1201
Qy 690 DTFITPVFKKIEKKEBEKNEPTFQVSKKKONPVNHSQLESHEKEDLQREB--HSQKS 746
Db 1202 SNSIPIPIIKIRKKNNDN--DNIRKNDNINSNNINSNDKCLFLSKERQRYVHLKNN 1258
Qy 747 DSTKQVATVLDKNNISSKSTNNPNK 773
Db 1259 NIIVNNTMMFRQSNSCDNTTSLKXK 1285

RESULT 5
118427
hypothetical protein C035c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18427
R:Lawson, D.; Bowman, S.; Barrell, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18427
A:Status: preliminary; translated from GB/EMBL/DBJ

```



```

Db      567 EETEREIGINTESLNTKTFER-----VKENVTNLIKIKELKHVYFSDGKGNITKY 618
Qy      481 -----DGDIMDGMK--DYKANGPDP-----VYTMDDGNV-----YLQYSDILNAK 520
Db      619 TDKIKCINDDITMAVSQOQIDQHINGLDDIOKSESYSVEKQEIINLEKYSNTEISNDVNE 678
Qy      521 AVGVHYOFLYDYNVKEVNI-----DPRKNTSIEYADKSVVFNNI-----560
Db      679 GIKKKQOQIIVTKIDKKNIYEBINKLSIKIEKNDTSLKVKQDINISYQGNLGNLFLE 738
Qy      561 --DKRANGPDGEIQOHYYIN-----GKEYTSFNDIKQIIDXTLAN-IKI-----601
Db      739 QIDBEKKKAEKNTIKSMKAYIIDLDNIIKKKSQEIETEMDIIKMDINKEMELKISHDDKKC 798
Qy      602 -----VYKDFARNTYVKEFLTNIDGTGSEVSLKPRVYVTLQNGKB 641
Db      799 HDKRNHKNKNTISDIYDKSSKTIIDPSRESIDND-IKKNLQKNVSSQNHNSDIN-----851
Qy      642 MSTIVSEEDPILPYKGELEKGYQPDGMBISGFEKGDAGYVNLASDTPFKPYFKLIE 701
Db      852 -----QCLAEVANIYNIILKLNKIKKIIDKVK 877
Qy      702 EKKEBENKPTDVSQKKNPQVNHSQLNESHKEDLQREHSQKSDST---KQV-----T 753
Db      878 EYTSIEIK-----NKNINDELNNSEKVIKKIIEGLSLKCRSKINSTLDDKDIDECIKN 932
Qy      754 ATVLKNNISSKSTNN 770
Db      933 INVLKKNILNEETITN 949

```

RESULT 7

228676
rhoptry protein - Plasmodium yoelii (fragment)

```

C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676, A45521
R:Sinha, K.A.; Keen, J.K.; Ogum, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: 220507; MUID:97077455; PMID:8920022
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q26216; EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AA8412
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, W.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KE>
A:Cross-references: GB:M34281

```

```

Query Match      5.7%; Score 230; DB 2; Length 2401;
Best local similarity 21.0%; Pred. No. 0.04;
Matches 187; Conservative 138; Mismatches 292; Indels 274; Gaps 46;

Qy      2 LGRIASRPNKMGKSGSLKKDITGV-HHQENESIKKSSFTIDRNISTIDFANK 60
Db      589 IDKIYTKFENILTDKTELETGFGLSLNHNSNN---KELIYFYDLKAKLGRKEN- 644
Qy      61 DLKLLIKKKPREVDFTSETGRMEYDYKYDDKNITLA-----YDGTGLEVET 110
Db      645 -----MLTIQFNEK-----KAVEDIKKKQVNDINKIYSNIEITTYTSITVNNBTEMEI 693
Qy      111 EK-----LDEIKSKTYGVLSPSKDGHEPILKISINVSNNAKVYVGN 151
Db      694 GKSLIPLNTYVLEKVKANVTMLNBIKEL-----KQYDQDPQK-----EKNIKYPPEN 742

```

```

Qy      152 NYKS-IEIKATKYDHFSTMTFEDLYAN---INDI---VDGLAFAGMRELKYQNDQKKA 203
Db      743 KIKNDIDTLNQGIDKSIETLTF-EIKKSENHIDEIKQIDIKKAKPKNKTF--NEDPKEI 799
Qy      204 EIKI-RMEBKIKETSEYPPVSYGNVLELGGD---LSKNDPLATJMES-GKIY-----254
Db      800 EKKIENIYEKIDKKNKIYKIDKLANESIKIENDTSLKELKNNINLSYKSLGNLPLQOI 859
Qy      255 -SDSEKQOYLAK-----DNI-----ILKGYA 275
Db      860 DEKKKKAHTIKAMAYIIDDLDNIKKKSQOIEKENINIMDIKMDIHKKMKALNISHDDK 919
Qy      276 LKVTI--YNPKDTMLKNGVYSKEDIKIQKAPNLALSETTIYADSRNVEDGRSTOS 333
Db      920 IYHTTSKHHBEKISDIRKNSLKIIDPFSESYINDIKKELEKRVLESQNNNTDINQ---975
Qy      334 VLMSLDC-FNIRQVTFPKNDGELALDDQGNLVTOSSKLYLFKQDKXEYGR-----387
Db      976 -YLSKIENTYNIIL-----KLNKIKKIIDK-----VKETTDIEKKN 1010
Qy      388 DEFNVE-----AIREDSGLPIDTKPVNLSDKNYFNPSSKNKIYVRNDEFTLRGK 438
Db      1011 KKNLNLSENSEKIITQLAKENSLSKESQSK-IKSTIDNVVSECIN---ITLKYTYI---1063
Qy      439 ISDKGFNWEKLVNESVVDNY-LIYGLAHIDNTRDPNKLANYGDGDMDWGKDYKANGF 497
Db      1064 VIEKKNINITYFNKAEYQNVSLNPNNTIEMADTKSQYI-LNIX-----KNNG- 1109
Qy      498 PDKVTMDGNYVLTQYGSGLNAKAGVHYQFLYDYNVKEVNIIDPRKNTSIEYADKSVYF 557
Db      1110 -----TNNTDINIKELKHK-----KSNVYDDEKGRTOELKKNKEL-- 1147
Qy      558 NINDKRNNGPDGEIQOHYYINGKEYTSFNDIKQIIDXTLAN-KIYVDF--ARNTYVKE 614
Db      1148 -----PEKYGEVYVLAN-KYAV--ELKKNFDTKQVYSEQIIEIKDPAHNTFTSQ 1195
Qy      615 FLANKDQSEVSLKXRYVTIYQNGKESSTIVSEEDPILPYKGELEKGYQPDGMEISG 674
Db      1196 --ADSEKKMEIKNEQRIEDEVAKNNKSN-----KALD-----1229
Qy      675 FEKKKADAGVNLASDTPFKPYFKLIE---KKEBENKPTDVSQKKNPOV-----723
Db      1230 -----IQLSVEEP-KIKFLKIKDLRTKSDCLKKTDIEKISNLSIDTQETKLI 1278
Qy      724 -NHSQLANSHRKEDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 773
Db      1279 ENKNILNLTLEKLSLESLKNGKKNIEDQKK-----LDEVNSKIKNIESNVNQ 1324

```

RESULT 8

118440
hypochetrical protein C0425W - malaria parasite (Plasmodium falciparum)

```

C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18440
R:Lawson, D.; Bowman, S.; Barrett, B.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z18935
A:Accession: T18440
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4550 <LAW>
A:Cross-references: UNIPROT:O77336; EMBL:Z98547; NID:e1325376; PID:e1325396; PIDN:CAB
A:Map position: 3
A>Note: C0425W

```

```

Query Match      5.7%; Score 228; DB 2; Length 4550;
Best local similarity 21.6%; Pred. No. 0.11;
Matches 193; Conservative 119; Mismatches 347; Indels 236; Gaps 43;

Qy      5 IASRFPKMLGNGKE---GSLKKDITGV-----HHQENESIKKSSFTIDRNISTIRD 56

```

```

Db      746 IEBSRISGSGNHEIYGASIKCAQNGVRNKLQGNHNDIKDKNKKOG---DKKDKNDKD 802
Qy      57 FENKQIKKLIKKEFREVDPTSETGKRMBEYDYTKDKNIAIYDDGTLEIE----- 109
      803 GQKQDDKKGDDKRYDDGDD-----KYDDDDKRYDDDDDI--YDDDDDFNFHDSSTSK 853
Qy      110 -----TEKLDEIK-SKIYGLSPSKDGHFELIGKIS-NVSNNAVYVYG 150
      854 RLSTNDLSLKQDGKSLKKEDETLKCSKKNINVOEKESKNMYCMANTNCCVARE-KNYFY 912
Qy      151 NNYKSIEIKATKYDFHASKTMTFDLYANINDIYDGLAFAGD-----MLTF-----VKDN 198
      913 NDYMHINNTLSDHEKMINNMKSPGNLPLFDVDFHFKQDFIMLNTLIKIKELIQVLYG 972
Qy      199 DQKAEIKIRMPBEIKETSEYVYSSYGNV-----ILAGSDUS-----KN 240
      973 NYKQVCVCFKKTGTGK-TYCMPEPSCSTYTHIYCYNRYMQUHVRKKKNDILCASQKSRK 1031
Qy      241 KPDNLTJYME-SGKITSDEKQOYLKQNIILKGYALKVTTNPGKTDMLBGNVYSKE 298
      1032 QRAASVYKRGAGSKQKQDGRSLDRKQDKVTRRSNKEYSANDSDGKN--NGDDKKKGD 1089
Qy      299 DIAKI QKANPRLALSETTIYADSRNVEDGRSTQSYLMSALDGFNIIRYQVPTFKMNDK 358
      1090 NIGKNNNNNNNNDDSDNNNGDDSNNGDDSNNG-----DGNNTINGDNNNINGDGN 1143
Qy      359 EALDDQGNLYT-----DSKLYLFRKQDEKTYGDEKFN 392
      1144 NNINDEGNNDNDCLEBKQGDQKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRK 1202
Qy      393 EAIKEDGSMFLIDTYKPNLSMDKNYFNPSPKSNKIYVRNPEFYLRGCI-SDKGFNMELY 451
      1203 ---KDH-----DQKDSPSKINNRSKN-----KVKDKVGN-----I 1234
Qy      452 NESVDNTIYIGDLHDITRDPTNLIKLVKQD---IMDKMKDYTFANGPDKVTMDGN 507
      1235 NDDKINN-----NTNDNNKNNNI--NGDNKNNNINGDNNKNNING--DKI-----NNN 1278
Qy      508 VYLQGYGLMAKAVGVH-QFLYONVKEVNIIDPKGNTSIEVADKSVVFNINDK----- 562
      1279 INGDKINNNTIDDKISKNYSHSNYDN--ELTHEQKNNFN-KNKTKRYKRSINDKEND 1335
Qy      563 ---RNGFDE-----IOEHIYINKEYS-----FNDIKQIIDKTANIKIVKDE 607
      1336 ILNNKNDPTDKLNTNFMLEYQKISSDKITSISNNKDIKNIMD-----IKDK 1387
Qy      608 RNTTYKPEILANKOTGEVSELKPHRYVTIYQNGKMSSTIVSEDFILPYKGELEKGYOF 667
      1388 DIKNIKDI---KDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDIKDI 1424
Qy      668 DQWELSGFEGKKDAGV-----INLSKDT--FLKPVFKIEE--KKEENKPTFDVSKKK 718
      1435 ---IKDIKDVNDVSVHLNCKVNNKSDTQDPCPSDDKINNIIRKKNKINIKTDQVY--- 1477
Qy      719 DNPQVNSQLNESHKEDLQREBSQKSD-STKDVATATVLDKNNISSTKSTNNPN 772
      1478 -NPLVDSSTSTDCNQCNYKQKKEKAVYKIGMINIKIKLHNNNNNNNNNNNNNNNNNN 1531

```

RESULT 9

T18429
 hypotheical protein C0345w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T18429
 R:Lawson, D.; Bowman, S.; Barrell, B.
 Submitted to the EMBL Data Library, August 1997
 A:Reference number: Z18935
 A:Accession: T18429
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1711 (LAW)
 A:Cross-references: UNIPROT:O77322; EMBL:Z89547; NID:e1325376; PID:e1325381; PIDN:CAB11J

C:Genetics:
 A:Note: PFC0345w

Query Match 5.5%; Score 221.5; DB 2; Length 1711;

Best Local Similarity 19.7%; Pred. No. 0.059;

Matches 160; Conservative 134; Mismatches 274; Indels 243; Gaps 37;

```

Qy      29 EHHQENESIEKES-----FTIDNISTIDPFENKDLK-----KLKKGFREVD 74
      984 DNNNDNNNNNNMYDSSSSNNHNYIILTNDRKLN-MDNFINNNLEINNNSQNVIEKNLEYIN 1042
Qy      75 DF-TSETGKRMBEYDYKQDKNIIAYDDGTLEIEKDEIKDEIKYGLVSPSKDGHFE 133
      1043 NVKLTITS-----NYQSNNTNSKDE-----NHISSTKSKEDPTLNTSRKSSYE 1086
Qy      134 ILGKI-SNVSQNAKVYGNMYKSIKIKATKYDFHASKTMTFDLYANINDIYDGLAFAGDM 191
      1087 YNNKILQSTNSKSLANGAYENNLPSGKKKQK-----GVLYKDI-BHINDIQD--KYPEDL 1138
Qy      192 RL-----FYKDDQKAEIKIRMPBEIKETSEYVYSS--YGNVIELQSDLSKRP 243
      1139 NINCVRKYVTENEKRL-----LPLEL-----EYNLVSSDEKFG-----LMTKIND 1179
Qy      244 NITRMSGKIYSDSEKQOYLKQNIILKGYALKVTTNPGKTDMLBGNVYSKEDIKI 303
      1180 N-----NITVMQGHQYHLYDNN--QKKHILEPTNK-----VSI 1212
Qy      304 QKANPRLALSETTIYADSRNVEDGRSTQSYLMSALDGFNIIRYQVPTFKMNDK--GEAI 361
      1213 QR-NNNINSVITKTHNEVEKKNKDKRN-----YDNFTCDKKKKIYNNII 1255
Qy      362 DQDGNLYTSSKLYVLRKQDEKTYG-----BDKFNVAIKEDGSMFLIDTYKPNL-- 411
      1256 NSDKDIYHNN-----IITTKNEKSGIIGNILNRDKDITNELLDVGVKXELDTFKOSYID 1312
Qy      412 ---SMQKNYFNPSPKSNKIYVRNPEFYLRGKISDQGFNMELYVRNPEFYVLYIGDLHDN 469
      1313 CHKCKENILMTNKKK-----EDHQIIDVAK-IFN---ETMTIMDNKKIYDDKAVHE 1362
Qy      470 TRDPNLIKLVKQGDIMDKMKDYKANGPDKTYTMD--GNVYLQGYSLDLMAKAVGVHGY 527
      1363 KCTTHNDVTHHMMDDILSTSIKNNBENLFDITYQOKRIRIDIYNNR----- 1407
Qy      528 FLYDNKPEVNIIDPKGNTSIEVADKSVVFNINDKNNNG---PDGEIOEHIYINKE 583
      1408 ---INILQEDDDDNHNN-----NNNNNNNNKILIFETKDDQDLHNNKNNL 1452
Qy      584 TSFNDIKQIIDKTANIKIVKDFARTYKPEFLANKOTGEVSELKPHRYVTIYQNGKMS 643
      1453 EGTBEFSDFIEKKNKIKIKKQKESYHKIDESLISNKKNNKVSIL-----LINNKDSS 1505
Qy      644 STIVSEEDFILPYKGELEKGYOFQDQWELSGFEGKKDAGVYINLSQDTIKPVFKIEEK 703
      1506 SV-----DNNKNNNNKNNNNNN-----NNNN 1527
Qy      704 KEENKPTFDVSKKKDNPQVNSQLNESHKED-----LQREBSQKSDTKDVATATV 756
      1528 NKKNNKNNNNDSFSDNNLINDNNNNNNNNDSFSDNNLINDNNNNNNNNNNNNNNNNNN 1587

```

RESULT 10

A24594
 probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 09-Jul-2004
 C:Accession: A24594
 R:Holder, A.A.; Lockyer, M.J.; Odink, K.G.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, Nature 317, 270-273, 1985
 A:Title: Primary structure of the precursor to the three major surface antigens of Plas

A:Reference number: A24594; MUID:86014355; PMID:2995820

A:Accession: A24594

A:Molecule type: DNA

A:Residues: 1-1640 <HOL>

A:Cross-references: UNIPROT:E04933

C:Superfamily: major merozoite surface antigen

C:Keywords: surface antigen

Query Match 5.4%; Score 219; DB 2; Length 1640;

Best Local Similarity 18.6%; Pred. No. 0.071; Mismatches 340; Indels 228; Gaps 39;

Matches 167; Conservative 164;

```

23 KDTTG-VEHHHQBESIK-----EKSFTIDRNISTIRDPENDLKK-----LKKK
232 KDNNGMEDYIKKNNKTENINELIESKTTIDKNNATKEE-----KKKLYQAY 283
76 FTSETGRMEEDYDYDKDNIIAYDDGTDLEYTEKLEDEIKSIKYGLVSPKDGFEIL 135
284 DLSTYKQLEB-----AHNLISV-----LE---KRIDTLK-----KRNKIKELL 319
136 GKISVSKAKVYVGNVNNYSIEIKATKYDFH-----SKMTPTDLYANINDIYDGLARA 188
320 DKINIEIKPPANSNGTPTLIDKNNKJIEHEKEIKELAKTIKFNIDSLPTDPL----- 373
189 GDMELFVNDNQKAKIKRMPKIKETSEYPP-----YVSXYGVN-----IELGE-GDLSK 239
374 -ELEAYLREKN-KNIDISAKVETKSTENEPNGVTYPLASINDINNALNELNSFGDLI- 430
240 NKPDNLTYMESGKIYDSERKQYL--LKNIDIL-----RKGYALKVTTYNPGKTDML 289
431 -NPFDTYKPSKNITLYDNEKKYINIEIKIKIKKKKESDKSYEDRSKSLN---DIT 485
290 EG-----NGVYSKE-----DLAKIQQA-----NPNLPALETTIYADSRVDEGSGTOSV 334
486 KEYEKLNEIYDSKFNNNIDLTNFERMMGKRYSYKVEKLTHTNTPTAFSYENSKNLEKLT 545
335 LMSALDGP---NI-----RYQVFTPKNDKGALIDKGNLVTDSKVLFGKDKEXT 385
546 ALKTMEDYSLNITVEKEKLYKNLISKIENIEITLVE--NIKDBEQLP-----EKKIT 598
386 GEDKFNVAIKEDGSMFLPIDYKPVNLSDMKYFNSKSKKIYVRPEFLRGKISDGKGF 445
599 KDNKPPDKILFVSDIVKVOQVLL--MNKIDELKKTQLILKAVE--LKNHIVPNYSY 653
446 NMELRVNVSVDNLYLYGDLHIDNTDFNPKL-----NVK----- 480
654 KQENKQOB---PYLLIVLKKGEIDKLVMPKVBLSLINEKKNIKTQGGSDNSEPSTGEI 709
481 -----DGDIMDGKMDYK-----ANGFPDPRV 501
710 TGOATTKGQAGSALBEGSVQAQAEQQAQPPVPVPEBAKAQVPTPPAPVNNKTEAV 769
502 TDMGNGVTLQGYSDINA-----KAVGHYQPLYNVAKPEVNIIDKGNISIRYADGSV 556
770 SKLD---VLEKLYQPLNTSYICHKYLIVSHSTNNEKILKQYKITVEESKSSCCPLDIL 826
557 FNIN-----DKRNGFPGS---EIOEGHIYINGKEYTSFNDIKQIIDDITLNIKIYV 603
827 FNIGNINIVNYSWMPISLNSLSQLEMEIYERKEMVCLYTLKNDKIKIKNLEBAKAYSVISV 886
604 KDPANNTVKEPILANKQGEVSELPRHVTVTIONGKEMSSITVSEBDFILPVYQGELEK 663
887 KTLSSSSWQPLSLTPQDPRVSAANDTSHSTINNSLKLFEINILS-----LGNKNG 937
664 GYQPDGMEISGFEK-----KQAGVYNLSKOTFIKPVFKIKBEKKEENKPTP--D 713
938 IYQ-----ELIGKSSSENFYKELKLDSDFTFNSFTNPFVASKADINDSLNDSKRRKLEED 993
714 VSKKDNQVWNSQLESKREKDLQREBSQKSDSTKVATATVLDKNNISSK-STTNP 771
994 INKLKKTQLSLPDLNKKYKLEKLEFLDKKKTIVGKTKMQIKKLTILKEQLESLNLSLNP 1052

```

RESULT 11

SAZOKI

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #ext_change 09-Jun-2000

C:Accession: A25120

R:MacKay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stuenkelberg, H.

EMBO J. 4, 3823-3829, 1985

A:Title: Polymorphism of the precursor for the major surface antigens of Plasmodium f

A:Reference number: A91030; MUID:86136024; PMID:3004972

A:Accession: A25120

A:Molecule type: DNA

A:Residues: 1-1631 <MAC>

C:Comment: The merozoite stages of different strains have strain-specific surface ant

C:Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The

C:Superfamily: major merozoite surface antigen

C:Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane pr

P:1-19/Domain: signal sequence #status predicted <Sig>

F:20-1631/Product: major merozoite surface antigen #status predicted <MAT>

F:67-84/Region: 3-residue repeats (S-G-T/P)

F:1614-1631/Domain: membrane anchor #status predicted <MBN>

F:97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (asn

Query Match 5.4%; Score 218; DB 1; Length 1631;

Best Local Similarity 18.1%; Pred. No. 0.078; Mismatches 339; Indels 240; Gaps 37;

Matches 164; Conservative 162;

```

23 KDTTG-VEHHHQBESIK-----EKSFTIDRNISTIRDPENDLKK-----LKKK 69
223 KDNNGMEDYIKKNNKTENINELIESKTTIDKNNATKEEKKLYQAYDGLFYNNKQ 282
70 FREVDFTSETGRMEEDYDYDKDNIIAYDDGTDLEYTEKLEDEIKSIKYGLVSPKX 129
283 LEBANILSVLEKRID-----TLKKNENIK----- 307
130 GHEPILKISVSKAKVYVGNVNNYSIEIKATKYDFH-----SKMTPTDLYANINDIY 182
308 ---ELIDKINEIKNPPANSNGTPTLIDKNNKJIEHEKEIKELAKTIKFNIDSLPTDPL 364
183 DGLAPAGDRLFVNDNQKAKIKRMPKIKETSEYPP-----YVSXYGVN-----IELGE 234
365 -----ELEAYLREKN-KNIDISAKVETKSTENEPNGVTYPLASINDINNALNELNS 416
235 -GDLKKNKPDNLTYMESGKIYDSERKQYL--LKNIDIL-----RKGYALKVTTYNP 283
417 PGDLI--NPFDTYKPSKNITLYDNEKKYINIEIKIKIKKKKESDKSYEDRSKSLN- 473
284 GKTDMLBEG-----NGVYSKE-----DLAKIQQA-----NPNLPALETTIYADSRVDEG 328
474 ---DITKEYEKLINEIYDSKFNNNIDLTNFERMMGKRYSYKVEKLTHTNTPTAFSYENSKN 530
329 RSTQSVLSALDGP---NI-----RYQVFTPKNDKGALIDKGNLVTDSKVLFGK 379
531 LEKLTALKTMEDYSLNITVEKEKLYKNLISKIENIEITLVE--NIKDBEQLP----- 584
380 DDKRYTGEDKFNVAIKEDGSMFLPIDYKPVNLSDMKYFNSKSKKIYVRPEFLRGK 439
585 -EKKITKDNKPPDKILFVSDIVKVOQVLL--MNKIDELKKTQLILKAVE--LKNH 638
440 SPKGFNMEELRVNVSVDNLYLYGDLHIDNTDFNPKL-----NVK----- 480
639 HVPNSYKQENKQOB---PYLLIVLKKGEIDKLVMPKVBLSLINEKKNIKTQGGSDNSEP 694
481 -----DGDIMDGKMDYK-----ANGFPDPRV 501
695 STEGEITGOATTKGQAGSALBEGSVQAQAEQQAQPPVPVPEBAKAQVPTPPAPV 754
496 GPPDYVTDMDGNGVTLQGYSDINA-----KAVGHYQPLYNVAKPEVNIIDKGNISIR 550
755 NKTENVSKLD---VLEKLYQPLNTSYICHKYLIVSHSTNNEKILKQYKITVEESKSS 811
551 DGKSVVFNIN-----DKRNGFPGS---EIOEGHIYINGKEYTSFNDIKQIIDDITL 597

```

Db 812 DPLDLLFNIGNNIPVMSFPLSNNSLSQLFMEIYKEMVNCNLKLDKNDKDKNLLBEAK 871
Qy 598 NIKVYVADPANTTYKBEPLKACOTGSEELKPHRYVTYTIQNGKMSSTIVSEEDPILPVY 657
Db 872 KVTSTSVKTLSSSSWQPLSLTPQDKPEVSANDTSHSTLNANSLKLFENILS----- 922
Qy 658 KGBLEKGYQFQDMEISGFEKG-----KDAGYVNLNSKOTFIKPVFKIEKKKEBENK 709
Db 923 LGKKNKITY-----ELIQKSSSENFYEKLIKDSPTFNESFTNFVSKADINDSLADESK 978
Qy 710 PTF--DVSKKQDPQVNSQLNESHKEDLQREHSQSDSTKDVATATVLDKNNISK-S 766
Db 979 KLEEDINMLKKTQLSLFQVYKYLKLERLPDKKTVGKYMQLKLTLLKEQLESKLN 1038
Qy 767 TTNP 771
Db 1039 SLNMP 1043

RESULT 12

major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (sc
S05603
N/Alternate names: gp195, surface antigen
C/Species: Plasmodium falciparum
C/Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C/Accession: S05603, S04850
R/Myer, P.J.
submitted to the EMBL Data Library, April 1989
A/Reference number: S05603
A/Accession: S05603
A/Molecule type: mRNA
A/Residues: 1-1639 <MYL>
A/Cross-references: UNIPROT:P04933; EMBL:X15063; NID:g9896; PIDN:CAJ3163.1; PID:g9897
R/Myer, P.J.
Nucleic Acids Res. 17, 5401, 1989
A/Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plas
A/Reference number: S04850; MUID:89345116; PMID:2668887
A/Accession: S04850
A/Molecule type: mRNA
A/Residues: 1504-1639 <MYL2>
A/Cross-references: EMBL:X15063
C/Superfamily: major merozoite surface antigen
C/Keywords: glycoprotein; merozoite; surface antigen
F/1-19/Domain: signal sequence #status predicted <Sig>
F/20-1639/Product: major merozoite surface antigen #status predicted <Mat>

Query Match 5.4%; Score 218; DB 2; Length 1639;

Best Local Similarity 18.6%; Pred. No. 0.079;

Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KDTTG-VEHHQENESIK-----EKSFTIDRNISTIRDPENKDKLKIKKGFREVD 75
Db 232 KDNVGRMDYIKKDKTIENTINELBESKTIIDKRNKKER-----KKGLYQAY 283
Qy 76 FTSETGKMEEDYDYDKNIIAYDDGDLLEYTEKLDIISKIYGVLSPEKDGFEIL 135
Db 284 DLSINRQLES-----AHNLSIV-----LR--KRIDYTLK-----KRNKIKEL 319
Qy 136 GKISVSNNAKYVGNNTKSIKIKATKYDF-----SKMTFQDLYANINDIYDGLAFA 188
Db 320 DCLNINIKPPRANSQTPNTLDDKNNKQIEHKEIKIYAKTIKFNIDSLFTPLP----- 373
Qy 189 GDMRLFVNDQDKAEIKIRMEKIKETSKRYP-----YVSRYGVN-----IBLGE-GDLSK 239
Db 374 -ELAYLBEKN-KNIDISAKVETKESTBENEPNGVTYPLSYNDINNANLAINSPGDL- 430
Qy 240 NKPDULTGTEGSKIKYSDEKQYL--LKDNIL-----AKGALAKTYYTTPGKTML 289
Db 431 -NPDYTEPSKNITYTDERKKFIBIEIKIKIEKKKIESDKKSYEDRKSLN---DIT 485
Qy 290 EG-----NGVSKS-----DIKIQKA-----NPNILASETTIYADSHVDEGSGTOSV 334
Db 486 KEYKELAIYIDYKSKNNNIDLTNFRKQKRSYKVEKLTTHNTYASYNSSGNLEKLT 545

Qy 335 LMSALDF-----NII-----RYVFTFGNDKGEALIDKQNLWTDSKYLFGKDKEXT 385
Db 546 AKTMEDSLNRITVVEKELKTKNLIISKIENIEFLVE--NKKQBEQLP-----EKKT 598
Qy 386 GEDKPEVNAIKEDGSMFLIDTFVNLSDMKYFNFSKSKNIYVRNPEFLRGKISDKGCF 445
Db 599 KQENKPDKILLESVIDIVAVQGVKVL-----MNRIDELKTKQLLKQVE--LKNHIVPNYS 653
Qy 446 NMELVNESVDNNTIYGDMLIDTFRNRIK-----NVK----- 480
Db 654 KOENKOE-----PYLILVLRKEIDKLKVPMPKVESLINEKKNIKTEGSDNSBPTEGHI 709
Qy 481 -----DDGIMQMKQDYK-----ANGPDKY 501
Db 710 TQOATTRGQAGSLBEGDSVQAQOEQKQAPPPVPPPAKAQVPTPPPAVNNKTAV 769
Qy 502 TMDGNVYLGQGYDLNA-----KAVGVHYOFLYDNVYKPEVNIIDPKGNTSIEYADKSV 556
Db 770 SKLD-----YLRKYBRLNTSYICHKYLIVSHSTNNEKILKQYKITBEESKLSGCPDL 826
Qy 557 FNIN-----DKRNGFDG---EIOEQHIYINGKAYTSFNDIKQIIDTKINIKIY 603
Db 827 FNIGNNIPVMSFPLSNNSLSQLFMEIYKEMVNCNLKLDKNDKDKNLLBEAKKVTSTV 886
Qy 604 KDFARNTTYKBEPLKACOTGSEELKPHRYVTYTIQNGKMSSTIVSEEDPILPVYKGELEK 663
Db 887 KTLSSSQWPLSLTPQDKPEVSANDTSHSTLNANSLKLFENILS-----LGKKN 937
Qy 664 GYQFDGWEISGFEKG-----KDAGYVNLNSKOTFIKPVFKIEKKKEBENKPTF--D 713
Db 938 IYQ-----ELIQKSSSENFYEKLIKDSPTFNESFTNFVSKADINDSLADESKRKLLED 993
Qy 714 VSKKQDPQVNSQLNESHKEDLQREHSQSDSTKDVATATVLDKNNISK-STTNP 771
Db 994 INKLKKTQLSLFQVYKYLKLERLPDKKTVGKYMQLKLTLLKEQLESKLNINP 1052

RESULT 13

lipoprotein (imported) - Mycoplasma pulmonis (strain UAB CT1P)
C/Species: Mycoplasma pulmonis
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C/Accession: G90603
R/Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samsen, D.; Gallison, F.; Moszer, I.
Nucleic Acids Res. 29, 2145-2153, 2001
A/Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pu
A/Reference number: A99512; MUID:21267165; PMID:11353084
A/Accession: G90603
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1546 <KUR>
A/Cross-references: UNIPROT:Q98P17; GB:AL445566; PID:g14090150; PIDN:CAC13908.1; GSPDB
A/experimental source: strain UAB CT1P
A/Genetic code: SGC3

Query Match 5.4%; Score 217.5; DB 2; Length 1546;

Best Local Similarity 20.2%; Pred. No. 0.077;

Matches 180; Conservative 125; Mismatches 336; Indels 249; Gaps 40;

Qy 1 KLGEIASKTKNIGNGKESLK-----KDTTGVEHHQENESIKESKSTFIDRNISTIRD 56
Db 103 KKEILASQVGTSVNNEKNFRILBAKOTKTKNDPNSAKSQWLDPSQDHLKNSQSKNIN 162
Qy 57 FENKDLKLIKKGREVDFTSRTGKMEEDYDYKD--DKGNIIAYDDGDLLEYTEKLD 115
Db 163 INNEOSKQLQTLANNEBSANLQTONILSEXTQDLDPN-----NSNKEKELKN 214
Qy 116 IKSKIYGVLSPEKDGFEILK-INSVSNNAKYVGNNTKSIKIKATKYDFSKTWTPTPL 174
Db 215 VLS--FNGQAVKDSFERYFYEKIVSKLESKINTDKKVDLTLDKTFADF-----SPHQ 267

QY 175 VANINDYDGLAFAGDMRLFYKNDQKKAIEIKIMPEKIKETSEKPYVSYGVNIEIGE 234
 DB 268 Y--IETIKOLFKRAST--IKDTYO-----TNKIFLTYDEY----- 238
 QY 235 GDLSSKRPDNLFTME-----SGKIYS-----DSKQOYLKKNIIILKRGVAL 276
 DB 299 --IKKNBSNKLKDKKPPSNPENEIYKPLEKDALVBIQALHKOYLEKNEIIKITGEIYD 356
 QY 277 KTTNPGKTMLENGG-VYSKEDIKTI-----OKANPNTALIS 314
 DB 357 KVAALFKSRBELSKAGALLFSKQOSAKITOLINHLISRYPPAPENILELSEKAGLINEL 416
 QY 315 ETTIYADSNVEDGSGT--QSVLMSALDGFNIIRYQVTFPMQNDKGEALDKOGLYDSSK 373
 DB 417 KEETIYKGEVBIAYFTLQALISNADQOKIFL-----NKNLKIDR--SIVVDKN-- 464
 QY 374 LVLFGDKDEKYTGED--KENVBAIKEDGSMFLPIDTKPVNLSMDKNYFNPB----- 421
 DB 465 ITIPAKSNVTITRKOSKSFPTMFITVQKALTFEIAEBSQSINLNGLSGTFKDESSLVKI 524
 QY 422 -KSNKTYRANPEPYLAGKISDKGFWELRVNBSYVDNTLIYGDHLIDNTRDPN----- 474
 DB 525 EKNALVLVAKTGTAFINSKSPSKYQ-----SVFEN--YGSVYIEGAKIWNVSESG 572
 QY 475 -----IKLNVKQDIDMDGKQYKANGPDK-VTDMGNYVLTQVYSDLNAKAVGVH 525
 DB 573 GIIRNHWGSLTPKNGEI-----RDNIISG--DKGIYISQGNALISGSDIKMS----- 620
 QY 526 YQFLVDNVRPEVNIIDPKGNTSIEYADKSVENINDR-----NNGPDGEIIOEQHIYING- 580
 DB 621 FSSILNLE-KTNINSGSIVNNAVSVKSLFEIDNSKIQISNNALINPGSSAIFLKN 679
 QY 581 -----KEYTSFNDIKQIIDKTNIKIV----- 602
 DB 600 STMHLAGSLKIKREASBORIEVLVLPFOAKLISPENIISLDNYQGLSALFKITSVKN 739
 QY 603 VQDPAR-----NTTVGEFILNKDT-----GEVSELKPRVTVTIQNGKMSSTIVSE-DF 652
 DB 740 INDFKVPFLVMTYKKEFFFLMDPTKLFPVNYKTLKQNHLLIOSGDESETEKIIKDELDF 799
 QY 653 -----ILPVYKSELEKGYQF-DGWEISGPEGK-----DAGVYINLSK 689
 DB 800 YRPTAAVVKLILTLQVLTIPKHKFWEBAFEYFNPFLQKWEYETINELIRIDPFYLEDIAY 859
 QY 690 DFFIKVPFKIEKKEBENKPTFVSKKQNDPQVNSQUNESHKEDLQR 739
 DB 860 PEFV-----ENGMLPEPBY-IHTNVNPFVLEH-----FRNEQVAR 894

RESULT 14

B42771
 reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
 C/Species: Plasmodium vivax
 C/Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #ext_change 09-Sep-1997
 C/Accession: B42771
 R/Galinaki, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
 Cell 69, 1213-1226, 1992
 A/Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
 A/Reference number: A42771; PMID:92315338; PMID:1617731
 A/Accession: B42771
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Cross-references: GB:M88098; NID:g160627; PID:g160628
 A/Experimental source: strain Belem, merozoites
 C/Genetics:
 A/Gene: RBP2

Query Match 5.4%; Score 215.5; DB 2; Length 1252;
 Best Local Similarity 20.9%; Pred. No. 0.071;
 Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;

QY 2 LGIASSKPK---NLGNGKEGSLKQDTTGVEHHEHSEBSIKERSPT-IDRNI----- 51
 DB 266 LSEIKKIKDKCTTETISNKG---KQIELEKPKPBEBSNKNVNIENENIRNBEQY 322
 QY 52 -----STIRDFENDLKLKIKK--FREVDPTS 78
 DB 323 LKQIEDAEKQSTKVELFHHGETTISNIFASELIGVETSKQKINABDIOMEIERHNS 382
 QY 79 EFKRMEEDYKTDKDNIIAYDGDYDLETEKLEIKETIKYGVLSPSQGHFELIGKI 138
 DB 383 EIQYVKGQDENLKNLNEPNNYDABD-ELANNKSTNAKYLIEITNLESVGN-----LSEI 437
 QY 139 SNVSKAKVYVGNYSYI--EIKATYDFSKTM-----TFDYAN-INDIVQCLAPAGD 190
 DB 438 TNLKQGEKTY-SKADIMKQIATSENTKTEKLEKVDQGSYVYVNLQITTT-----B 490
 QY 191 MRLFYDND-----OKBAIKIRMEKIKK-----TK-----S 218
 DB 491 RNLITVEKRLNGIDSTITINIEGALKESKGNVEIGLEKLEIEIGKRNKLKVDITKKSINS 550
 QY 219 EYPPVSYGVNIEIGEDLSKNPDLTTHESKITYSDSEKQOYLKKNIIILKRGVALY 278
 DB 551 TVGNFSSLFNNPDLNQYDFRNKINDYENKN--GEIYNFEGSLNKISENL--RNASENT 605
 QY 279 TTYNPKETDMLBEGNGVYSKEDI--AKIQANPMLALSETTIYADSRNVEDSGSTQSVL 335
 DB 606 SDYNSAKITRLER--QKEKVLNKEBEANIKYLR-----DYKATSEPR----- 646
 QY 336 MSALDGFNIIRYQVTFPMQNDKGEALDKOGLNLTSSKVLVFGDKDEKYTGEDKPVNEAI 395
 DB 647 -----FLFNMK--ESLDKINEMIKKQLTY-----NKGNGVQLVYENI 683
 QY 396 KEDGSMFLPIDTKPVNS-----MDKN-----YFPPSKNKIYANPEPYLAGKI 439
 DB 684 KE-----LVDEN--NLSDLIKQATGKBEIQTITHTLAKKAKATILGHVPTSAKYGIKI 736
 QY 440 SDRGGRFWBLRVNESVVD-----NYLIYGDHLIDNTRDPNIKLVNVD-- 481
 DB 737 TP-----BALTELLGDALKTAQELKFPESKNNVLETENMKNTELDVHNRKIDQAYK 790
 QY 482 -----GDMDWKKQDYKANGPDDKVTDMQGNVTL-----QTGYSDLANAKAVGVH 525
 DB 791 VALEIILASHDEIDTKQD-----SSKLIEMNQIYLKVLINQYRNKISSIKSEAVS 844
 QY 526 YQFLVNVK-----PEVNIDPKGNTSIEYADKSVENI-----NDRKNGPDGEIOEQ 574
 DB 845 VKI--GNVSKKSELSKITCSYDNIITALEKQTELQNLNRSFTQEKTYTNSDKLE-- 900
 QY 575 HIYNGKEYTSFNDIKQIIDKTNIKI-----VVQD-- 605
 DB 901 -----KIKTDFESLKNAL-KTLBGEVNALKASSDNHHEHVQSKBPVNPALSEIEKETD 953
 QY 606 -PARNTTVGEFILNKDTGVSILKPRVTVT--IQNGKMSSTIVSEBDITLVYKGELE 662
 DB 954 IDSLNALDELLKKGRTCEVSRYKLKDVYTKESIDTBLINTLEKN--VAAYLAVIK 1009
 QY 663 KGYQFQWELISGPEBGAGVYVINSKDTFIEKV-----FKKIEBKEEENKPTFD-- 713
 DB 1010 KATV-----DTYQD--VLTINHEPTKQVSHHEPTNFPKSKSSSELTKAVYDSK 1057
 QY 714 -VSKKK-----DNPQVN-----HSQUNESHR-KEDIQREBSHQ 744
 DB 1058 TIISKIKGLVILEVENTBMNTLESSAKIEALYNEILKNKXTSLNEIYQTSNEVYLOEMKS 1117
 QY 745 KSDSTQDYTA--TVVD--KNNSISKSTTN 770
 DB 1118 NADKIYDSKIEFTVLDTKSKSNVTYQHSINN 1149

RESULT 15

A45597
 DNA-directed RNA polymerase (BC 2.7.7.6) III largest chain - malaria parasite (Plasmo
 C/Species: Plasmodium falciparum

